

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE				
1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY NA		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE NA				
4. PERFORMING ORGANIZATION REPORT NUMBER(S) IDA Paper P-1945		5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION Institute for Defense Analyses		6b. OFFICE SYMBOL (If applicable)		7a. NAME OF MONITORING ORGANIZATION DoD IDA Management Office
6c. ADDRESS (City, State, and Zip Code) 1801 N. Beauregard Street Alexandria, VA 22311		7b. ADDRESS (CITY, STATE, AND ZIP CODE) 1801 N. Beauregard Street Alexandria, VA 22311		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION OUSDRE (R&AT)		8b. OFFICE SYMBOL (If applicable)		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA 903 84 C 0031
8c. ADDRESS (City, State, and Zip Code) The Pentagon Washington, DC 20301		10. SOURCE OF FUNDING NUMBERS		
		PROGRAM ELEMENT	PROJECT NO.	TASK NO. T-5-310 WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) Skill Acquisition Curves and Military Training				
12. PERSONAL AUTHOR(S) Norman E. Lane, Essex Corporation				
13. TYPE OF REPORT Final	13b. TIME COVERED FROM 3-85 TO 1-86		14. DATE OF REPORT (Year, Month, Day) January 1986	15. PAGE COUNT 192
16. SUPPLEMENTARY NOTATION				
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) skill acquisition, learning curves, training	
FIELD	GROUP	SUB-GROUP		
19. ABSTRACT (Continue on reverse if necessary and identify by block number) This paper summarizes currently available information on the major factors that influence the rates at which individuals acquire new skills and knowledge when they are being trained.				
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Jesse Orlansky			22b. TELEPHONE (Include Area Code) (703) 845-2293	22c. OFFICE SYMBOL

DD FORM 1473. 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

IDA PAPER P-1945

SKILL ACQUISITION CURVES AND
MILITARY TRAINING

LIBRARY
RESEARCH REPORTS DIVISION
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA 93940

✓ Norman E. Lane
Essex Corporation

January 1986

Prepared for
Office of the Under Secretary of Defense for Research and Engineering



INSTITUTE FOR DEFENSE ANALYSES
1801 N. Beauregard Street, Alexandria, Virginia 22311

IDA PAPER P-1945

SKILL ACQUISITION CURVES AND MILITARY TRAINING

Norman E. Lane
Essex Corporation

January 1986



INSTITUTE FOR DEFENSE ANALYSES

Contract MDA 903 84 C 0031
Task T-5-310

EXECUTIVE SUMMARY

The learning curve is a well-known description of the fact that the rate at which an individual improves his skills and knowledge decreases as training continues. Too little training can cause problems later in retaining and using previously acquired skills and knowledge. Too prolonged training raises skills and knowledge only slightly and thus is a waste of resources. Therefore, it would be useful to understand how to establish an optimum point beyond which further training yields diminishing returns in learning, according to various types of tasks, difficulty of the tasks, ability level of students, the proficiency required at the end of the course and the proficiency required some time later on the job.

This paper summarizes and evaluates information relevant to these issues. Four forms of learning curves (power function, negative exponential, hyperbolic and logistic) are described and evaluated by being applied to 15 sets of learning data. In this limited test, all forms could fit the data reasonably well; goodness-of-fit (R^2) ranged from 0.74 to 0.99; 39 of the 60 fits were 0.90 or higher. The power law provides one of the best fits and is consistent with current cognitive theory about basic processes underlying task performance. The logistic fit is excellent in a purely statistical sense but it offers no immediate interpretation for the processes underlying learning. The negative exponential form, which has a broad base of support in the literature, yielded the least appropriate fit.

Nevertheless, what is known about learning curves cannot, at present, be used to determine the optimum length of a training course, although a path in that direction can be seen. A useful start needs data on individual student progress curves for various segments of typical courses. Such data appear to be available although, in collecting it, we need to distinguish between individual and group progress curves and to identify significant segments within courses to which learning curves could be applied. At the same time, we should collect data about student ability and the difficulty level of various course segments. It is important to know that this type of information can now be collected routinely and without interference in courses that use computer-based instruction.

ACKNOWLEDGEMENTS*

The ideas of many individuals appear in this document. Dr. Jesse Orlansky made major contributions to formulation of the problem and provided a number of substantive suggestions. Dr. Martin Tolcott, Dr. Marshall Farr and Dr. Richard Gibson also made important content and editorial inputs. Appreciation is also due to Dr. Marshall B. Jones for sharing his thoughts on the subject matter in several discussions and to Dr. Jack Adams for generously providing a copy of a draft monograph highly relevant to the subject matter of this report. It is likely that none of these individuals would agree with all the content of the report, but each was instrumental in shaping some portion.

*This study was prepared by Essex Corporation, 1040 Woodcock Road, Orlando, Florida, under subcontract to the Institute for Defense Analyses. The work was performed for the Office of the Deputy Under Secretary of Defense for Research and Engineering (Research and Advanced Technology), under the technical cognizance of Captain Paul R. Chatelier, the Military Assistant for Training and Personnel Technology.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES AND FIGURES	viii
LIST OF ABBREVIATIONS	ix
TECHNICAL SUMMARY	S-1
INTRODUCTION	1
BACKGROUND AND REQUIREMENT	1
Characteristics of the Military Training Environment..	1
Constraints in the Military Environment	4
Finding Mechanisms for Improvement	6
Requirement	7
OBJECTIVES	9
Acquisition	10
Retention	10
Linkage between Acquisition and Retention	11
DOMAIN OF ANALYSES	12
Scope and Emphasis	12
Comments on the Military-Relevant Literature	12
APPROACH	15
INITIAL DIRECTIONS	15
MODIFYING APPROACH AND EMPHASIS	17
EMPHASIZING RETENTION AND TRANSFER	17
EMPHASIZING PROCESSES OF ACQUISITION	20
ACQUISITION CURVES, SHAPES AND PARAMETERS	23
BASIC PARAMETERS	24
THE NATURE OF ACQUISITION FUNCTIONS	26
Generalized Power Function	28
Generalized Exponential	30
Hyperbolic	32
Logistic	33
Linearity Transforms	34
Positive vs. Negative Acceleration	36

COMPARATIVE ANALYSIS OF ACQUISITION FUNCTIONS	38
The "Universality" of Learning Functions	40
An Information Processing Viewpoint	41
The Power Law	43
Exponential Equations	45
Contrasting the Functions	47
Inconclusiveness of the literature	47
Some additional data	48
CONDITIONS THAT AFFECT CURVE SHAPE	55
Nature of the Task	56
Task Difficulty	59
Degree of Prior Learning	62
Plateaus	64
Criteria for Termination (Mastery Level)	67
Mastery training	71
Distribution of Practice	74
Interference and forgetting	76
Individual Differences	76
Different Training Methods	78
GROUP VS. INDIVIDUAL CURVES	79
Characteristics of Group Curves	79
Characteristics of Individual Curves	80
Problems in Generalization	80
Factors in Discrepancies	82
Resolving the Controversy	83
BEHAVIORAL VS. ENGINEERING APPROACHES	84
Nature of Tasks Used	85
Emphasis on Group Output	86
Very Long Time Periods	86
Motivation and Interest of Trainee	86
Handling of Poor Performers	87
SOME RELEVANT THEORY AND FINDINGS ON ACQUISITION	90
THE NATURE OF SKILLED PERFORMANCE	93
DISTINCTIONS BETWEEN SKILLS AND ABILITIES	95
Fixed-Abilities, Changing Tasks Models	96
Changing Abilities Models	99
Implications	100
STAGES AND PHASES OF LEARNING	101
Fitts' Stages	102
Anderson's Stages	104
Rasmussen's Paradigm	106

DEVELOPMENT AND IMPLICATIONS OF "AUTOMATIC" BEHAVIOR	108
SCHEMA THEORY EXPLANATIONS OF ACQUISITION	112
CONTEXTUAL INTERFERENCE	116
KNOWLEDGE OF RESULTS AND FEEDBACK	119
PART-TASK AND TASK SIMPLIFICATION APPROACHES	124
PREDICTION OF TIME TO TRAIN	127
RETENTION	129
ISSUES IN TASK AND SKILL CLASSIFICATION	130
Categories of Classification Systems	131
Categories of Taxonomies	132
GENERALIZATIONS AND DISCUSSION	134
OVERVIEW	134
GENERALIZATIONS FROM THE DATA	135
Interpretation and Uses of Learning Curves	135
More Emphasis on Acquisition than Retention	137
Selection of Termination Points for Training	137
Procedural Tasks	138
Impact and Management of Individual Differences	139
Stages of Skill Acquisition	140
Simplified Part-Task Approaches	140
Contextual Interference	141
GENERALIZATIONS ABOUT THE DATA	141
Descriptive vs. Comparative Data	141
Theory vs. Applications in Studies and Data	142
Constraints on Implementation	145
Sources and Uses of Historical Data	145
Instructional Systems Development and Classification Systems	150
Training Effectiveness Data	151
CONCLUSIONS	153
RECOMMENDATIONS	155
OVERVIEW	155

1. COLLECTION AND USE OF TRAINING DATA	155
2. MORE FLEXIBILITY IN TRAINING TIME AND SCHEDULING	156
3. PROGRAMS FOR REFRESHER TRAINING	156
4. R&D ON TASK AND SKILL DESCRIPTION SYSTEMS	157
REFERENCES	159

LIST OF TABLES

Table 1. Goodness-of-Fit of Several Learning Functions to Selected Data Sets	50
---	----

LIST OF FIGURES

Figure 1. Parameters of Typical Acquisition Curves	27
Figure 2. Curves of Positive, Negative and Changing Acceleration	37
Figure 3. Power, Hyperbolic, Exponential and Log (Trials) Fits to Star Tracing Data	52
Figure 4. The Effects of Task Difficulty -- Successes in Ring Tosses from Varying Distances	61

LIST OF ABBREVIATIONS

AFQT	Armed Forces Qualification Test
ASVAB	Armed Services Vocational Aptitude Battery
ATI	Aptitude-treatment interaction
CI	Contextual interference
CMI	Computer-Managed Instruction
CTA	Critical Task Analysis
DTIC	Defense Technical Information Center
IP	Information processing
ISD	Instructional Systems Development
KR	Knowledge of results
KP	Knowledge of performance
ROI	Return on investment
SKA	Skills and Knowledge Analysis
TSL	Time spent learning
TTL	Time to learn

TECHNICAL SUMMARY

INTRODUCTION AND BACKGROUND

One of the most difficult problems in military training is establishing the "appropriate" length of time for a training course or module. Too little training for too short a period causes problems later in remembering and using skills and information in additional training or on the job. Training past some "optimum" point gives diminishing returns in learning and wastes resources. Trainees learn at different rates, and each is likely to be at a different level of learning when training is terminated after a fixed time period. It would thus be useful to be able to estimate the "best" time course of training (length, schedule and intensity), given the task to be trained, an entering trainee group of some typical ability level, the proficiency needed by the end of the course, and the proficiency required at some time after the course is completed.

Manipulation of time-course variables in military training appears to offer considerable leverage in improved learning and in more efficient resource management. Although decisions about course length, content and minimum performance standards are made routinely throughout the training delivery system, these tend to be "intuitive" in nature rather than "data based." Overall, little or no formal attention has been given in training decisions to the multitude of variables which influence the amount of time needed for training, and their effects are not well understood. This report reviews and synthesizes data and findings from the literature on skill acquisition, learning, retention and transfer which deal with the domain of time as it influences performance, i. e., with the changes in capability to perform that occur as a function of time and practice. The emphasis is on assessing the applicability of available information to the estimation of training time course and to

related areas of potential training improvements. (Note that the terms "acquisition" and "retention" are used here as indicating the learning and remembering of skills and knowledge, not in the other commonly encountered sense of personnel acquisition and retention.

OBJECTIVES

The overall objective of the present analysis was to look for "lawful" relationships between characteristics of the training environment and method of training and the degree of useful "learning" attained at various periods in the training process. In particular, the focus was on variables which affect training time-course in a positive or negative way, and on the influence of these variables on training performance, retention, and transfer. The analysis is considered as an initial step in determining a) whether or not it is worthwhile to modify duration, scheduling and intensity of courses and segments to achieve more effective training, b) if so, whether existing data are sufficient for identifying and making decisions about needed modifications, and if not, c) what data and information are required and how they should be obtained and used.

SCOPE AND EMPHASIS

Selection of literature for review and analysis was heavily biased toward studies most applicable to military training situations. An extremely broad literature on learning, training, skill acquisition and retention, transfer and forgetting bears directly or indirectly on the main thrust of the review. Except when there were substantive theoretical implications in interpretation of findings, preference for in-depth analysis was given to studies with task content, training context, stimulus and response requirements and subject populations which were most closely analogous to and most representative of military tasks and environments.

A further delimiter concerned a focus on acquisition studies with implications for retention or transfer. The viewpoint used throughout this analysis was that the goal of a training segment was to provide a capability to do a job proficiently or to perform well in later training. Greater emphasis was placed on studies which dealt with performance in post-training contexts as well as within training or learning segments.

APPROACH

The primary concern in analysis of relevant studies was the examination of changes in performance of a task across time, reflected in what is typically referred to as an "acquisition curve." The initial phase focused on determining if published research on acquisition and learning provided evidence of learning curves (or rate data) which would support quantitative predictions of how much time a unit of training should require. The general approach was a) to locate studies reporting learning curves or rates and their associated parameters, b) to look for variables or characteristics of the learning situation which might cause curves to differ on shape and level parameters, c) to attempt to extract regularities in the behavior of curves that were sufficiently reliable to be useful for prediction of time courses across a variety of training situations, d) where possible, to quantify effects of key variables on rate or terminal performance, and e) to link findings to current theoretical developments on "processes" underlying skill acquisition to derive recommendations for modifying existing approaches to training.

Outcomes of initial analyses indicated that: a) While there were substantial regularities in the shape of acquisition curves, their parameters varied widely, and the emphasis on quantitative prediction was not supportable from existing data; b) the literature itself was insufficient to fully verify or reject the possibility of such prediction. While learning curves tended toward a common negatively accelerated form, no

usable patterns were apparent in the parameters. The literature was primarily descriptive in nature rather than comparative, the majority of tasks used were too simple or otherwise not representative of military jobs, and key information for use in a quantitative framework was often omitted. While considerable information from acquisition studies could be brought to bear on the military training question, best use of that information required a modified approach. Emphasis changed from the primary focus on curves and rates as an end-product to their use as a means of extracting reliable statements about the influence of key variables on the outputs of military training and education that could be generalized across training situations.

FINDINGS

The Military Training Environment

The military training environment is a difficult arena in which to introduce and evaluate training innovations. The report describes some characteristics which tend to make military training unique, among them an unusually high turnover of personnel in the system, a wide variability in content and difficulty among military tasks, similar variation in the entry-level abilities of trainees and a requirement for geographic and organizational separation of training sites. These define a series of constraints on the ways in which training can be conducted. Training is typically organized into segments or modules administered sequentially, with learning built up over segments. The need for a steady, predictable output from training tends to cause reliance on fixed course lengths and group training. These and other "inherent" constraints on training efficiency are discussed and their impact elaborated.

Acquisition Curves and Functions

In a description of acquisition curves, the basic parameters, equations and mathematical properties of

commonly-used acquisition functions are presented. Various forms of the power function, exponential, hyperbolic and logistic curves are described and related to their bases in the acquisition and learning literature. Conditions that can cause variations in curve shape are identified, and their implications for interpretation of data discussed. Differences between group and individual performance curves that affect generalization of group data are defined. The literature on appropriate use of group curves is controversial; arguments for and against generalization are summarized, and a resolution to the controversy is suggested.

There has been extensive use of "learning curves" (also called manufacturing progress functions) for decision making in industrial settings. The engineering literature on production functions is outlined and compared to the behavioral literature on learning and skill acquisition. It is concluded that the tasks used and the improvements in performance with experience in industrial settings are more applicable to the military training environment than those typically found in research on learning in the behavioral science laboratories.

Comparing Acquisition Functions

The typical curve relating performance to time or practice is a negatively accelerated function in which the gain in performance on a trial decreases as practice continues. Two major function families -- power and exponential -- have been suggested as the "typical" shape of the acquisition function. There is significant support in the literature for both types of curves. The power law, referred to by its advocates as "ubiquitous," is somewhat more substantiated by data and has been more extensively studied. The power law function takes a linear form when both performance and time are in logarithmic units (the "log/log linear" fit). Developments by cognitive theorists suggest the power function as the logical form of task

data derived from a variety of mathematical assumptions about the nature of basic processes underlying task performance.

To develop additional evidence on the form of acquisition curves, some data from the literature were fitted with several of the common functions. Results supported the general superiority of the power law and that of a simpler equation, logarithmic only in trials (semi-log/linear). It was concluded that there is no one functional form likely to be "best" in all circumstances. Conditions were described for which different curve shapes are likely to be encountered, but the overall robustness of the power law was noted.

Theoretical Issues

Some current theories on skill and knowledge acquisition are reviewed as they relate to the effects of time and practice on performance changes. In general, the literature is consistent with respect to the effects on acquisition performance of most variables likely to be manipulated in military training, particularly the time and practice variables emphasized in this analysis. There are factors of interest, however, that may affect training performance without a similar effect on learning. Variables such as distribution of practice, task difficulty and augmented feedback during training can cause increases or decreases in training performance with no effect, or an opposite effect, on retention or transfer. To consider manipulation of these variables during training, it is necessary to understand the nature of processes underlying their influences on skill acquisition and retention behavior.

The tasks performed by military personnel are typically characterized as requiring "skilled performance." Most recent research on skilled performance departs from earlier experimental work (that used nonsense syllables or reaction time tasks) in a concern for tasks which more closely approximate

those performed in the "real world." Tasks of this type have certain key characteristics. They require integration of multiple types of "skills" or components (motor, perceptual, procedural, etc.); they require extensive practice to gain a minimum capability; proficiency continues to develop almost indefinitely, and performance shifts, with continued practice, to an apparently "automatic" control mode. These and other aspects of skilled task performance are discussed and the theoretical implications elaborated.

Distinctions have been made in the literature between skills and abilities as separable aspects of an individual's capability to perform. Abilities have been considered as relatively "fixed" attributes, and skills as "learned" capabilities. The appropriateness of these distinctions has been questioned in more recent work. Evidence on this controversy and its treatment by current theory is presented. It is demonstrated that the distinction is unnecessary for time-course manipulation so long as trainee entry level is considered in training decisions.

A brief presentation of stage models of acquisition is given, which relates three similar views, derived from different theoretical positions, of what occurs within learning stages. Cognitive theory interpretations are mapped into those based on more conventional learning viewpoints. The development of "automaticity" of task control processes in late practice is discussed. Several theoretical structures are described which arrive at similar predictions about automatic behavior from divergent assumptions about the processes underlying skill acquisition.

The implications of "schema theory" for acquisition and retention are analyzed. The relationships of "schema" to acquisition behavior in a cognitive theory framework are described, and schema-based considerations for structuring

training content and for determining pacing and phasing of training are developed. The phenomenon of "contextual interference" is identified, in which manipulations of task difficulty produce decreased training performance but enhanced retention and transfer. The occurrence of contextual interference reinforces the need in training design to distinguish between training and post-training performance as indices of training effectiveness. It also has important implications for "guidance" training. Tasks trained under a single learning strategy may not be retained as well as those for which trainees can explore alternate ways of doing the task.

In additional discussions, current views on the role of knowledge of results, feedback and augmented feedback in improving learning are presented. Task decomposition and task simplification and their roles in complex skill training are discussed and related to a variety of part-task and simplified task approaches. Emphasis is placed on the importance of "cueing" and cue fidelity as factors in less than whole-task practice, particularly in simulators.

There is a small literature related directly to the issue of predicting the time required for training and the separation of trainees into subsets based on information about initial ability. This literature is summarized. While this approach may offer advantages in certain training situations, strategies for forming subgroups are likely to be highly situation-specific and to require both considerable information on entrant level ability and extensive historical data from prior training in order to be effectively implemented.

A major difficulty in estimating training time variables from task characteristics and trainee capabilities is the lack of good frameworks for describing tasks and skills. The impact of the lack of classification systems on the analysis of training requirements is described. Some issues in development of classifications and taxonomies for training are presented,

and it is concluded that the absence of consensual schemes is a critical deficiency in virtually all analytic or data-related approaches to improving training.

CONCLUSIONS RELEVANT TO MILITARY TRAINING

1. The "typical" curve relating training performance to practice has a characteristic negatively accelerated shape. Curves deviate frequently from that common shape, but usually as the result of one or more well-understood task characteristics or training conditions. The shape is sufficiently regular to form a "baseline" or target curve for training progress; major deviations from the general shape may represent undetected and/or undesired aspects of a training program that produce inefficiencies.

2. While curve shape (general form) can often be anticipated reliably, the time-course over which acquisition runs, and thus the curve parameters, is generally not predictable from prior knowledge of task characteristics. The mathematical description of a task learning curve requires data specific to a task or training segment.

3. There is much useful information in the learning and acquisition literature that could be applied to improve military training. The most valuable of these principles concern time-based aspects of training -- sequencing, scheduling, pacing and course length -- rather than training content per se. With a few exceptions, the benefit of these principles has not been realized in military training situations.

4. The main constraints on use of time-based principles for making decisions in military training are:

a. The fixed-time orientation in military training arising from the need to predict the availability of personnel for assignment.

b. Available data from the learning and training literature are not always sufficiently task-specific for generalization to military tasks.

c. Available data tend to deal with the practice of complete tasks, and do not generalize well to the cumulative acquisition of skills across training segments typical of military training programs.

5. Understanding the nature of learning which occurs on a specific training task as practice continues is critically important, both for efficient training and for avoiding undesired negative transfer effects. Determining the point at which "sufficient" training has been given (i. e., course or segment length) is complicated by a) a lack of task-specific retention and transfer data and b) the difficulty of distinguishing task conditions which increase actual levels of learning from those which improve training performance without enhancing learning. A number of recent developments in the theory of skill acquisition have potential for resolving the latter difficulty.

RECOMMENDATIONS

The following recommendations are in general restricted to those which have direct applicability to the structure of military training. Other recommendations, not directly germane to the main theme of time course estimation and manipulation, may be found throughout the paper.

1. Collection and Use of Training Data

Establish mechanisms in military training for the routine collection, maintenance, analysis and application of progress and performance data. Data should generally be collected and maintained at the segment or course level. Mechanisms can be

separate from or supplemental to any existing computer-managed instruction, but should produce data sufficient for a) defining typical progress curves for a course segment, b) identifying students who are having difficulties, c) determining appropriate course and segment lengths for both classes and individuals, and d) conducting formal and informal evaluations of training effectiveness.

2. More Flexibility in Training Time and Scheduling

Develop specific mechanisms which enable the adjusting of training time for classes and individuals on the basis of performance-related indices. Training decision-makers should have the flexibility to provide extra time, to rearrange schedules, and to vary the method and pacing of instruction as required to exercise quality control over the training product.

3. Programs for Refresher Training

Develop formal programs for routine provision of update and refresher training. These would initially focus on critical skills for specialists in selected jobs, with a gradual transition to all major job components for all specialties. Requirements for refresher training should be established by policy, and training should be provided on a regular basis, either in an operational setting or through consolidated facilities at a higher organizational level.

4. Research and Development on Task and Skill Description Systems

Develop a basic, standard notational system for describing military task requirements and trainee and operator capabilities.

INTRODUCTION

BACKGROUND AND REQUIREMENT

Military training is big business. The old truism that "half the people in the military spend half their time training the other half" is only a mild hyperbole. Each year, about 200,000 people spend about 20 billion dollars to educate and train the continuing input of new personnel and to upgrade the job capabilities of those already in the system. .

This paper will synthesize some data and findings bearing on military education and training and (as the data warrant) suggest some ways to improve current practices. Changes to the military system, however, even minor ones, are rarely straightforward. Some special characteristics of the military environment place real and practical constraints on training and education innovation and make it a particularly difficult arena in which to introduce and evaluate new ways of training. At the outset, we will consider some of these characteristics of the military training environment that influence the analyses and recommendations which follow.

Characteristics of the Military Training Environment

The nature of military training and the massive investment of personnel and dollar resources it needs each year is a direct outgrowth of the environment which generates these training and education needs and in which training must be conducted. The military training environment is in many ways unique among all classes of training and education. Some key characteristics and properties contribute to this uniqueness:

a. High levels of planned and unplanned turnover of personnel. By its nature, the military system incurs a steady loss of trained and experienced people through retirement and

routine turnover at the conclusion of enlistments. About 20 percent of all personnel each year, the combined numerical losses from the "top" and the "middle" of the system, are replaced by a continuing input of recruits at the "bottom." It requires an ongoing training effort to provide the basic set of skills and knowledge to the new recruits who will eventually migrate upwards and also leave the system in turn. As more experienced personnel exit, those still in the system must receive advanced training to move up and replace them, and the process continues with a cyclical upward spiral more extreme than in any civilian application.

b. Emphasis on training vs. education. A traditional distinction is drawn between education, which conveys a broad background in general skills and knowledge, and training, dealing with more task-oriented goals and objectives and more specific desired outcomes. Education is typically pointed toward a general preparation for further academic progress, training toward the skills required to do a particular job proficiently. Military systems represent a mixture of situations. Although much training, particularly in the earlier stages, approximates traditional education in structure and in its use of classroom and lecture formats, its content and course objectives are considerably more task-specific, and the predominant focus in most training situations is on direct skills training. The overwhelmingly practical emphasis of military training distinguishes it from other large-scale educational systems.

c. Wide variation in training task content and difficulty. The tasks toward which military training is oriented range from the conveying of simple factual information, to extremely straightforward procedural operations of only a few steps, to some of the most complex and highly integrated activities ever undertaken by humans. The total spectrum of military tasks imposes every conceivable combination of informational,

perceptual, motor, and cognitive requirements, and employs somewhere in the delivery system virtually every known variety of training approach. There are many different occupational specialties for which military training must provide preparation, each sharing common skill and knowledge elements with some other specialties. To take advantage of commonalities and to reduce training on irrelevant material, much military training tends to structure training for a job into a series of sequential course "modules" or "segments" of fixed content, with required skills and knowledge provided through appropriate combinations of these modules. The capability of trainees thus accumulates across exposure to a number of successive units.

d. Extreme variability of initial skills and ability in the entrant population. At virtually every stage of training, the military system must deal with a wide range of ability and a distinct heterogeneity of prior experience, and it must do so in most cases without being able to track or subdivide trainees on the basis of ability. It must, in other words, take a widely diverse raw material and bring each as close as possible to a standard level of knowledge and skills using the same training system, within a narrowly prescribed period of time.

e. Geographic and organizational dispersion of training. Inherent to the military structure that has evolved over time is a tendency, based on operational needs, to organize military units by functions and to separate those functions geographically. This can over time result in the segmentation of training into sequential "packages" keyed to organization and location and not necessarily to the best logical arrangement of material. Although efforts to consolidate training have reduced dispersion and some of the need for such sequencing, restructuring to avoid its impact has not always occurred.

Constraints in the Military Environment

The unusual properties of the military training situation impose in turn a number of constraints on the ways in which training can be conducted and bring into play factors which act and interact to inhibit some of the more effective approaches to enhancing learning and retention of critical knowledge and skills. Among these factors are:

a. The need to provide a steady supply of trained personnel to operating units in appropriate quantities at predictable times exerts a major influence on the structure and pacing of training. Courses other than fixed-content, fixed-duration ("lockstep" pacing) pose difficulties in timely assignment of people completing a course or training segment to their operating units or to the next training segment. This pressure toward predictable completion dates and lockstep scheduling has significant implications for training management and for the selection of training approaches.

In particular, the complexities associated with the varying of training time for individuals tend to inhibit or preclude the use of training strategies which may be more effective for retention of skill than those presently used. Later sections discuss, for example, concepts such as "guidance" training vs. "discovery" training. In guidance-oriented instruction, trainees are shown each step and provided with a single recommended problem-solving strategy. In discovery approaches, trainees are given direction toward solutions but encouraged to find their own strategies in addition to the single strategy or procedure provided in conventional training. There is evidence that individuals trained under discovery conditions both retain information longer and are better able to generalize it to other situations. Such "novel" approaches, which may increase training time in the short run (for at least some individuals) and decrease it in the total sequence, are difficult to

introduce and manage in a schedule established on the basis of fixed time per module.

b. The traditional organization of military training into segments and modules which results from the special characteristics of the military environment creates several difficulties. Under such a structure, it is important that each block of instruction provide the "enabling" skills and knowledge which allow for successful completion of the next module and build progressively toward the ultimate desired "competencies." An absence of appropriate enabling skills is cumulative. If the ordering of modules and their successive contents are not properly interrelated, trainees will fall farther and farther behind. This tendency to "lose the thread" of a course sequence and lag behind is often aggravated by the pressures created by fixed time for training and the emphasis on moving along through training segments. For trainees who are borderline but apparently coping, lack of time for "overlearning" basic fundamentals can still be acting to inhibit retention of learning information and its transfer to a later segment.

Segmentation imposes a further complication on retention. Delays can be encountered between successive modules and segments, causing loss of critical skills. Because there are likely to be many tracks composed of generic modules, there may be segments intervening between learning and use, and considerable time may elapse between the acquisition of skills and knowledge and the next segment or module in which those skills are needed. Further, the generic nature of course content and timing creates difficulty in providing for additional time required to refresh skills after periods of non-use. Such modularity and segmentation, while it is a natural outgrowth of military organizational structures, can have adverse affects on both acquisition and retention of important skills and information.

c. Many of the decisions about training approaches, strategies and course content in military training must be made without information on how likely it is that a given course or course segment will accomplish its objectives. Determining the effectiveness of military training is unusually difficult. Despite the considerable volume of output from military courses, the numbers of people with comparable training patterns is relatively small, often too small to form an acceptable evaluation group. Following training, members of this group become rapidly dispersed geographically, and each acquires an idiosyncratic accumulation of on-the-job experience that further reduces comparability. Under such circumstances, evaluation of training effectiveness using on-job criteria requires considerable effort and is extremely costly.

Finding Mechanisms for Improvement

As the preceding sections suggest, there is much in the structure of military training that is not readily susceptible to simple and rapid fixes of problems and inefficiencies. The inherent nature of training in the military environment can make application of conventional techniques for training improvement impractical or excessively expensive. The interlocking of modules and segments causes even simple changes in one portion of the system to echo through related segments and, if not carefully handled, to disrupt other parts of the system out of proportion to the magnitude of the change.

Any systematic approach to finding implementable solutions for some of the well-documented ills of the military training system must focus first on the leverage available in areas that are most readily under the control of training managers and decision makers. There are two aspects of the training situation that can be varied without major disruption of the predictability of course output. These are the overall duration of a course or segment and the individual effort required of a

trainee during the time course of training. Increasing the length of a course for all participants, while it adds a constant to total time required in the system, does not materially change knowledge about the availability of graduates for assignment or further training. Requiring additional hours and effort from a trainee within a course for remedial or enrichment purposes is done routinely in both military and conventional training, and is likely to be even more effective in the military setting.

There are many avenues and approaches through which current delivery of military training can be enhanced. Throughout the analyses, discussions and recommendations of later sections, a distinct (but not exclusive) emphasis will be on finding approaches that offer both real opportunity for training improvement and a practical potential for implementation under the constraints described above.

Requirement

Central to the improvement of military training is the need for better ways of selecting efficient and effective instructional methods and approaches which are appropriate to a particular training situation and, as noted previously, also offer some hope for actually being used in that situation. One of the leverage areas discussed above was variation of the duration of a module or segment. It is thus important to be able to estimate the "best" time course of training given the nature of the task to be trained, the characteristics of the entering trainee population, the ultimate proficiency needed on completion of the course, and the proficiency required after some elapsed time interval. Estimation of an ideal time course (the time actually required for learning) is heavily dependent on rate information, both the rate at which skills and knowledge are acquired and the rate at which they are forgotten.

Rate data describing changes in performance or level of mastery of material as a function of time or practice are clearly linked to time required to achieve a specified proficiency. Rate information, implicit or explicit, is central to determination of course duration. There is evidence in the current structure of military training that some implicit estimates of rate must already be in use. Each existing course or segment already involves some estimate of how long that course or segment should be, and incorporates in almost every case a provision for relating an individual's progress to "where he should be" at that point in training to identify trainees who are having problems. Some go further in establishing one or more explicit mechanisms for diagnosis and remediation of problems in the form of progress checks, additional instruction, repeating a segment, etc. All these existing provisions suggest that some "intuitive" conception of a rate of learning appropriate to a given set of skills and knowledge objectives is already embedded in most training situations.

These "embedded" estimates appear also to apply to questions of retention or transfer of learned skills. Informal feedback from later to earlier stages, however imperfect, apparently provides some indication of whether existing course duration and structure are sufficient to serve as a basis for later training or operational job performance.

The principal requirement, for which the present analysis is an initial step, is to find improved ways (preferably quantitative in nature) of projecting or estimating those rates and durations described above for explicit use in decision making about training. There is a second key component of decisions other than rates, durations and their associated efficiencies -- the cost of implementing changes compared to the resultant savings. Although cost considerations will be identified wherever possible, the development of cost relationships is not the purpose of this analysis.

OBJECTIVES

The broad objective of the present analysis is to examine what is known about skill and knowledge acquisition for the purpose of discovering "lawful" relationships between characteristics of the training situation, environment and method, and the degree to which the outcomes of training remain available for use when required on the job (retention) or assist in the learning of other skills (transfer).

The essential long-range goal of training is to provide the capability required to do the job proficiently and dependably, across continuing time. Thus the "proof of the pudding" for a training delivery system is how well people do their jobs later, often much later, after the conclusion of training. In many respects, the status or capability of an individual at the end of training is of little or no interest unless that status also has explicit meaning for later job proficiency. Training should bring about relatively permanent additions to the trainee's skills and knowledge base which are subsequently relevant to job performance or useful for learning other more complex material. We will at several points in discussion make distinctions between improved performance and improved learning. Some variables cause both learning and performance to increase and also thereby improve retention. Others bring about a performance improvement during acquisition but fail to enhance learning and have no desirable post-training effect. A key objective (and delimiter) of the present effort is thus to focus on acquisition that enhances retention, and where possible, to identify data from studies about learning or acquisition that translate into meaningful statements about retention, transfer or on-the-job demonstration of ability.

Efforts addressing the objectives are organized around the three major thrusts or sub-objectives outlined below:

1. Acquisition.

a. Examine data on acquisition rate during training with a focus on the shape of the acquisition "curves" which describe changes in skilled performance with additional training and practice.

b. Examine curve parameters (shape and level) as a function of type of material learned, method of training, amount and nature of practice, and degree of "mastery" or proficiency required (criteria for termination, overlearning, etc.).

c. Identify variables which systematically (and reliably) affect acquisition rate and/or training performance in ways which enhance retention (or transfer, as appropriate). The goal is the isolation of "families" or groups of curves which can be organized on the basis of factors above to forecast appropriate time course of training (how long should it take to learn skill S to criterion level L[2] by method M given entry level L[1]).

d. Analyze individual differences in acquisition (group vs. individual learning curves) and the importance of these differences in making generalizable statements about acquisition and retention performance and in decisions about course structure and length.

Emphasis is given to variables and conditions influencing acquisition and learning phases; concern with retention and transfer is strong but is predominantly contextual, used to define the domains of interest for analysis and to identify areas of particular emphasis.

2. Retention. Deals with examination and synthesis of data on retention of skills and knowledge after acquisition.

a. Identify key variables enhancing or reducing retention, primarily those which are active during acquisition (type of task or material to be learned, criterion or mastery level used, degree of overlearning, etc.), and secondarily those which help to maintain proficiency during periods of non-use.

b. Evaluate tradeoffs between investment of resources during acquisition vs. investment after training is completed. These include, among others, the degree of mastery or overlearning required in training to attain a desired level of retention of skills and knowledge, compared to equivalent expenditure of time/resources on refresher/update training and proficiency testing on the job.

3. Linkage of acquisition to retention. As noted previously, the goal of training is not just to improve training performance per se. The intent is to bring about changes which ultimately improve proficiency on the job. Selection of training approaches must consider the effects of changes and interventions on both aspects of performance, with retention and transfer perhaps the more critical of these variables. Linkage of the two phases is through the search for instructional approaches and strategies and training structure modifications which contribute to both rapid, efficient learning and to retention of acquired skills/knowledges.

Information in following sections relative to these objectives will not necessarily be found in any particular order. Discussions of acquisition will generally be interleaved with those on retention, and linkages between the two are likely to be scattered wherever they most naturally occur. Although some theoretical aspects of both will be discussed, predominant emphasis is on what can be done during acquisition to enhance retention; for in-depth descriptions of the processes underlying retention, the reader is referred to the more complete efforts by Farr (1986) and other writers to be cited later.

DOMAIN OF ANALYSES

Scope and Emphasis

Although an extremely broad cross-section of the literature on learning, acquisition, retention, forgetting, transfer, and so forth bears in one way or another on the objectives of this effort, the emphasis in selection of literature is heavily on the degree of generalizability to military training situations. Except when there may be substantive theoretical implications involved in the use of data for decision making (and there are a surprising number of such cases), preference in in-depth analysis was given to studies with task content, training context, stimulus and response requirements and populations which were most closely analogous and most directly generalizable to military tasks and environments.

Comments on the Military-Relevant Literature

With the exception of a small body of DOD-sponsored research, acquisition and retention studies with a clear and direct relevance to military training are extremely rare. The most recent Annual Review of Psychology article on training (Wexley, 1984) is notably lacking in citations dealing with military-related situations. Wexley deals primarily with private sector training, and notes the absence of academic research on topics such as computer-based instruction and other technologies particularly germane to military requirements. A strong emphasis on education (vice training) and classroom settings is also seen in previous Annual Review articles on instructional psychology (Gagne & Dick, 1983; Glaser & Resnick, 1972; McKeachie, 1974; Resnick, 1981). The majority of work on classroom instruction, particularly until very recent years, has been theory oriented and rarely tested against alternatives. Bruner (1966) described the sharp dichotomies then present between research that approached acquisition of skills from a

descriptive basis (learning theories) as opposed to a prescriptive basis (instructional theories). Glaser (1962) makes a similar distinction, pointing out the unfruitful separation of educational theories from experimental studies of learning, but has more recently (Glaser & Resnick, 1972; Glaser, 1982) noted the coming together of the two approaches. Gagne and Dick (1983) also comment on the gradual convergence of instructional technology and the psychology of learning into a common field with emphasis on cognitive factors in knowledge and skill acquisition.

Despite these encouraging trends, most of the kinds of tasks characteristic of military training have been addressed only superficially in the research literature. The performance of military jobs involves many different kinds of skills. The job-specific orientation of military training thus presents clusters of multi-component tasks, for which directly applicable studies of acquisition are particularly absent. Availability of relevant data is an even greater handicap for analysis of retention. In addition to a preponderance of work concerned with retention of verbal and informational material, the time periods employed are rarely representative of military tasks. As Glaser (1982) notes, most studies of retention (and acquisition) occur over conveniently short time frames, and don't examine the longer-term (months and years) periods associated with real-life competence.

A further divergence of the literature lies in its propensity to deal with tasks which are sufficiently self contained to be manageable within an experimental paradigm. These are typically complete tasks which are practiced repetitively as a unit. Acquisition in military training, as we have noted, is accumulative in nature, building as trainees move through the segments and modules of instruction. The meanings of "practice" in these contexts and the generalizations from studies of practice are subtly different for military training than for the research literature.

As later sections will show, translation of much of the acquisition and retention literature into a form that bears on the main thrusts of this effort will involve an unusually heavy reliance on theory to relate findings of typical studies to the somewhat different universe represented by military training tasks. The theories will of necessity be eclectic. McKeachie (1974) lucidly describes the shifts in the role of theory in learning. As the domain of experiments performed broadened from the laboratory into more applied settings, the well-accepted "laws" of learning no longer worked. All the established principles on knowledge of results, feedback, practice and incremental learning turned out not to be true at least some of the time. Statements derived from controlled laboratory studies often failed to generalize to behavior in situations where key variables interact in an uncontrolled manner or are constrained (as in military settings) by policy, doctrine or equipment effects in ways that prevent the expected outcomes. In such a context, theory becomes a tool for identifying such constraints and projecting probable outcomes if constraints could somehow be removed by changes in the system.

Sections that follow describe the approach used in the analysis of acquisition and retention literature, present some of the metrics of rate variables and the effects of training situations on these metrics, summarize the available data on acquisition and retention and its meaning for military training, and identify the uses of data for decision making in training. Discussions of each of these topics are, for convenience, presented separately for acquisition and for retention, although the crossovers between the two areas will be identified as they occur.

APPROACH

INITIAL DIRECTIONS

The initial phase of acquisition analysis focused on determining if published research on acquisition and learning provided evidence of learning curves (or rate data) which support quantitative predictions of how much time a unit of training should require. The general approach was a) to locate, in the literature, studies reporting curves or rates and their associated parameters, b) to look for variables or characteristics of the learning situation which might cause curves to differ on shape and level parameters, c) to attempt to extract regularities in the behavior of curves that are sufficiently reliable to be useful for prediction of time courses across a variety of training situations, and d) where possible, to quantify effects of key variables on rate or terminal performance..

Literature was identified through a variety of sources. Searches of DOD-sponsored work were carried out through the Defense Technical Information Center (DTIC). Open literature citations were acquired through DIALOG. Initial sources were augmented by personal libraries of the investigators and their colleagues, and identification of additional relevant materials continued throughout the effort based on directions indicated by initial reviews.

During the early part of analysis, it became apparent that the literature would not yield the systematic data and regularities required for generalization of effects and quantitative time course predictions. In particular:

- a. Most group acquisition curves had one of two general shapes, either logarithmic (power function) or semi-logarithmic (exponential), including variants with one or more plateaus

depending on (among other factors) task complexity, but parameters tended to be task specific and were resistant to ordering or clustering.

b. Key information required for interpretation of findings and of curve shape was often omitted. Entry-level abilities and/or prior experience, for example, were typically unexamined or unreported.

c. Almost without exception, individual acquisition curves, where reported, departed in varying degrees from the group pattern. This is neither a new nor a surprising finding. Hayes (1953) and Estes (1956), among others, have warned against generalizations from group curves to individual learning patterns. Both the above authors, along with Baloff and Becker (1967) and Hayes and Pereboom (1959), describe conditions under which the group curve may not be representative of any of the individual curves, and may both obscure and misrepresent the basic underlying form of skill acquisition. It is nonetheless true that designing courses or establishing course duration requires representations of group performance, since group curves define expectations for group members across a time period. This paradox is in part resolvable by a closer attention to the definition of "mastery" or the criterion of "completed learning" employed in a study. Group curves then become recast in the form of percentage of the group attaining the criterion as a function of trials or time. Such a resolution of the paradox allows (at least in theory) group data to be used for time course decision making, but requires data on variation in performance within the group and a realistic criterion level at which training can be terminated. Virtually no data were reported that met these requirements.

d. In order to maintain control of the experiment and work within a convenient time frame, many of the studies in the learning literature used tasks that were of insufficient

complexity or too purely "motor" or "verbal" to be representative of military training tasks. While results of these studies can be useful in the present effort, they must be considered as "indicative" of trends or "confirmatory" of findings in more generalizable task situations, rather than as directly applicable.

MODIFYING APPROACH AND EMPHASIS

Outcomes of initial analyses indicated that the hoped-for "families" of acquisition curves and the emphasis on quantitative prediction were unwarranted based on available data. No patterns of curve shape, rate, or level parameters were apparent. While considerable information could be brought to bear on the military training question, best use of that information required a shift in approach and overall emphasis. The global objective remained much the same, finding ways of using what is known about key variables in skill acquisition and retention to do a more effective job of training. Emphasis changed from the primary focus on curves and rates as an end-product to their use as a means of extracting from the literature reliable statements about key variables in military training and education which could be generalized across training situations. The main organizational schema for the effort thus became the detailing of information on how to train (or what actions a training manager could take) to reduce the decay of skills and knowledge across time.

EMPHASIZING RETENTION AND TRANSFER

The changes in focus described above require an explicit examination of studies from the standpoint of whether and how conditions of learning contribute to improved retention and transfer, not just to improved training performance in and of itself. We have noted previously the importance of capability to perform a job as the ultimate criterion of training

effectiveness. Retention (and to a lesser extent, transfer) is the principal intermediate criterion of successful training. The relationships between acquisition performance and performance after an interval of non-use (retention) or in a related task (transfer) are not always straightforward. Changes in training performance do not necessarily indicate concomitant changes in learning. These sometimes intricate interplays and the findings supporting their interpretation are addressed in detail in the literature discussion section, but are summarized briefly below without supportive citations.

a. Some training approaches have a positive effect on retention simply because they allow the emergence of a higher level of training performance and learning. In general, higher performance during training is typically associated with higher absolute levels of proficiency after elapsed time. These approaches contain no features specifically keyed to retention, but enhance it predominantly through the mechanism of higher performance (and presumably learning) at the end of training. High training performance is not always associated with better transfer to similar tasks; the "correct" amount of practice (and thus the appropriate duration of training) is a crucial variable in transfer. Too much training time results in inefficient training but rarely a loss of retention; too much practice and the associated overlearning of task-specific skills can actually reduce transfer.

b. Other approaches may improve retention (and transfer) without improving training performance and may even reduce apparent levels of proficiency during training. These approaches most typically involve presentation of tasks under unusually cue-rich conditions which make learning more difficult but provide many "anchors" or retrieval cues to assist in remembering the task or using its component skills in related situations. Because they focus on such areas as component learning or systematic practice of individual skills, they can

make distinct differences in the quality and retention of future learning. Processes believed to be operating to produce these somewhat anomalous outcomes are variously referred to as "contextual interference" or "schema formation," and the theory behind them is expanded in a later discussion section.

c. There are also approaches which materially increase performance during training but provide no enhancement of learning and hence no improvement in retention or transfer. These may involve such concepts as augmented cueing or feedback, special displays or other means of summarizing or presenting to the trainee task-related information not normally available within the task itself. Augmentation in training is a complex issue, with its outcomes heavily dependent on the success of the augmentation in promoting the development of insights, response patterns or motor programs that survive the removal of the augmented information. Naively employed, augmented training produces transient improvements during presence of the additional cueing, with no lasting benefit on training performance, and a potential negative effect on retention or transfer. Some recent work, with awareness of the limitations of the method, has given deliberate attention to acquiring greater understanding of the important elements in task cueing, and has shown much more promising results with respect to transfer. The theoretical implications of augmented cueing for acquisition of complex skills are considerable, and are likewise expanded in later sections.

As the above summary suggests, it is difficult within a single training situation to determine a priori if a particular increase in training performance which results from some change in the training situation is a "real" increase, as in (a) above, or a "transient" one, as in (c). Likewise, a decrease could be due to contextual interference as in (b), which would provide better retention, or to a change in conditions which simply produced less learning. Unless each change, modification or

intervention in training conditions is followed routinely by retention or transfer follow-up studies, it is necessary to look more closely at the content of the task and context of the training conditions, and to explore in a systematic manner the nature of what is being learned and how, to focus on the processes involved in skill and knowledge acquisition.

EMPHASIZING PROCESSES OF ACQUISITION

Over the last decade, it has become clearer that virtually all real-world tasks, military tasks in particular, are multi-component in nature. Tasks which can be legitimately viewed as purely motor or verbal or cognitive to the exclusion of other components have, if they ever existed, virtually disappeared from the military job structure. Continued successful performance of jobs involves the learning of several different, only partly related, job skills and their integration into a smooth, well practiced execution of a complete job unit which transcends any of its components. The emphasis on integration, planning and organization which even the simplest job entails involves learning of a type which has come to be labelled as "acquisition of cognitive skills." It is now widely recognized that virtually all non-artificial tasks, regardless of their apparent unidimensionality, require some form of strategy selection and evaluation and some planning components which are best described as "cognitive" in nature.

Gagne and Dick (1983), in an extensive summary of recent and ongoing work related to instructional methods, conclude that the field of instructional technology has become virtually indistinguishable in direction from the area of cognitive psychology, and has begun to share over time a steadily larger common base of concepts and terminology. This shift is in the most part due to the need to substitute for the "crumbling laws of learning" (McKeachie, 1974), which were quantitative and descriptive in nature, an improved understanding of how skills

are acquired, and, more importantly, how they are integrated into the kinds of skilled performance observed in real-life job situations.

As a later section will argue, there are two general directions that further studies of acquisition could take. One is to obtain data within naturally existing course structures, without change or intervention, and perform observational follow-up studies to examine the future success of trainees in later training and on the job. A second is to develop a program of proposed changes in and interventions into the current system which have potential for improved retention/transfer and to systematically evaluate both the changes and the efficacy of the concepts or theories which generated those alternatives. It is likely that both are potentially fruitful avenues. The latter direction, that of systematic modification, is obviously of greater complexity and materially higher risk. To present any chance of successful contribution to military training efficiency, work toward that direction will require a consensual focus on total job performance as the basic unit of analysis. Examination of variables such as extent of practice, time to learn, and organization of course segments and materials must be oriented toward that basic unit (or at least toward the intermediate criteria of retention and transfer). Because of the emphasis on integration of skills associated with this approach, it must of necessity be conducted within what current terminology would describe as a "cognitive learning" framework. Although many studies relevant to the present analysis take a much more elemental approach, discussions in later sections will attempt to tie together these elements in the context of ultimate relevance to job performance.

A considerable literature has been devoted to the description and analysis of functions relating time, trials or other practice indices to changing performance on a task. These studies in general show both a comforting regularity in the

shapes of acquisition curves and some interesting and revealing departures from these regularities. Understanding the forms that curves can take under varying learning conditions is an important element in developing ways of isolating factors that can enhance training performance and associated retention. The following section presents some of the equations descriptive of "learning curves" and identifies conditions which influence the shapes of these functions.

ACQUISITION CURVES, SHAPES, AND PARAMETERS

There are a number of practical reasons for knowing shape, level and other parameter information about acquisition functions. Among these are:

a. Estimating ultimate proficiency in a skill (eventual asymptote) for groups trained in different ways or for which it is not desired or possible to continue training for long periods. Conversely, one could estimate how long it would take to attain a given asymptotic level.

b. Determining when sufficient training or practice has been provided to achieve a desired level of performance known to yield a desired level of retention. This would include examination of the acquisition curve to detect the point at which "leveling off" occurs and "overlearning" starts. (Recent data from Jones (1985) suggests that the shape, particularly slope, of the acquisition curve at the termination of acquisition is an important indicator of retention). Application of this sort is likely to be more beneficial in the individual case than for the group.

c. Tracking individual performance to decide if a trainee is having trouble keeping pace with the group and whether intervention in training (and what sort) might be required.

The search for generalizable functions which characterize acquisition performance is by no means new. Most earlier work in the field involved the ultimate goal of universal theories or laws of learning. Analyses looked for shape regularities in order to derive theoretical implications or to determine consistency of empirically-derived functions with predictions from theory. Almost 70 years ago, Thurstone (1919) derived and tested a series of "learning functions" (see also Lewis, 1949).

In the 1930's and later, Hull (1943, 1952) proposed a number of mathematical laws of learning which included explicit predictions about the shape of acquisition functions.

The noting of regularities in the shapes of acquisition curves and their tendencies to follow a limited number of functional forms also began at least 60 years ago. Fitts (1964) credits Snoddy (1926) with the first identification of the "power law" form (log/log linear) of perceptual-motor learning curves, and there have been regular attempts ever since either to isolate the curve or to conclude that there is no such thing. As following paragraphs will show, both positions are in part correct. There are certainly recurring regularities in the shape of group learning (and possibly in individual learning) curves that hold across a variety of task domains. There are also curves that depart from these regularities. For both types of functions, the associated parameters have thus far been stubbornly task-specific. Further, the analyses conducted here, along with most others, suggest that the forecast of curve shape and level based on characteristics of a "new" task or body of material to be learned is not likely to be successful. As the section on using curves for making decisions about training will show, however, such inability to predict time course variables does not obviate the practical utility of curve data for managing training, particularly when historical data are available on tasks closely related to those for which estimation is desired.

BASIC PARAMETERS

Functions describing acquisition performance almost without exception can be viewed in terms of an ordinate (dependent variable - Y) measured in performance units and an abscissa (independent variable - X) cast in units of time, trials or other index of amount of practice or experience. Ordinates may be either in original units or in logarithmic transforms. The

ordinate will in general take one of two forms: a) Time to perform a given task (produce a fixed unit of output), or error scores, both of which produce a decreasing function across trials or practice, or b) output (number of tasks accomplished, units produced, performance score or rating) per fixed unit of time, which is an increasing function across practice. While these quantities are normally the inverse of one another, they create slightly different meanings for some parameters, and these distinctions will be noted in later descriptions.

To compare the various describing functions proposed by different authors, it is desirable to use a common notational scheme and to have clear definitions of parameters involved in the functions. The notation below will be followed as closely as possible during the following discussions (exceptions will be noted as they occur).

N = Number of trials, time units elapsed or other index of cumulative exposure or practice.

T = Time to respond, to perform a task, to complete one unit of output on a given trial, or other measure on which improved performance is indicated by decreasing values. Has an implied subscript of N indicating trial number.

Y = Performance score, output (units completed per one unit of time), ratings, or other measure on which improved performance is indicated by increasing values. Has an implied subscript indicating trial number.

A = Asymptote. The best possible level of performance. The value (variously) of T or Y as N approaches infinity. Sometimes used as the obtained value of performance at completion or termination of training or practice,

sometimes as theoretical limit attainable (to be fit from data). Can also be a preselected criterion or "mastery level" at which training is formally terminated.

B = Performance on first trial or output on first trial. May or may not be zero for performance or infinity for time per trial depending on E (prior learning). In a different notation, would be T_1 or Y_1 respectively.

E = Prior learning (in trial equivalents). Reflects transfer from prior experience or learning in terms of trials required to attain a presumed entry-level performance. Performance capability at N_0 . Related to the Y or T intercept (Y_0 , T_0).

R = Rate variable describing amount of change in Y or T with one unit change in N. Parameters referred to as "rates" are treated inconsistently throughout the literature. The meaning used here is of rate as the "average" slope of the curve.

Figure 1 illustrates the meaning of these parameters for both increasing and decreasing performance curves.

THE NATURE OF ACQUISITION FUNCTIONS

Most proposed equations (and most empirical ones) for describing performance changes during learning involve one of three major classes of functions: The power function, the exponential or the hyperbolic (which is a special case of the general power function). For certain measures of performance (particularly ratings), a fourth, the logistic function, appears to provide satisfactory fits to a variety of empirical data.

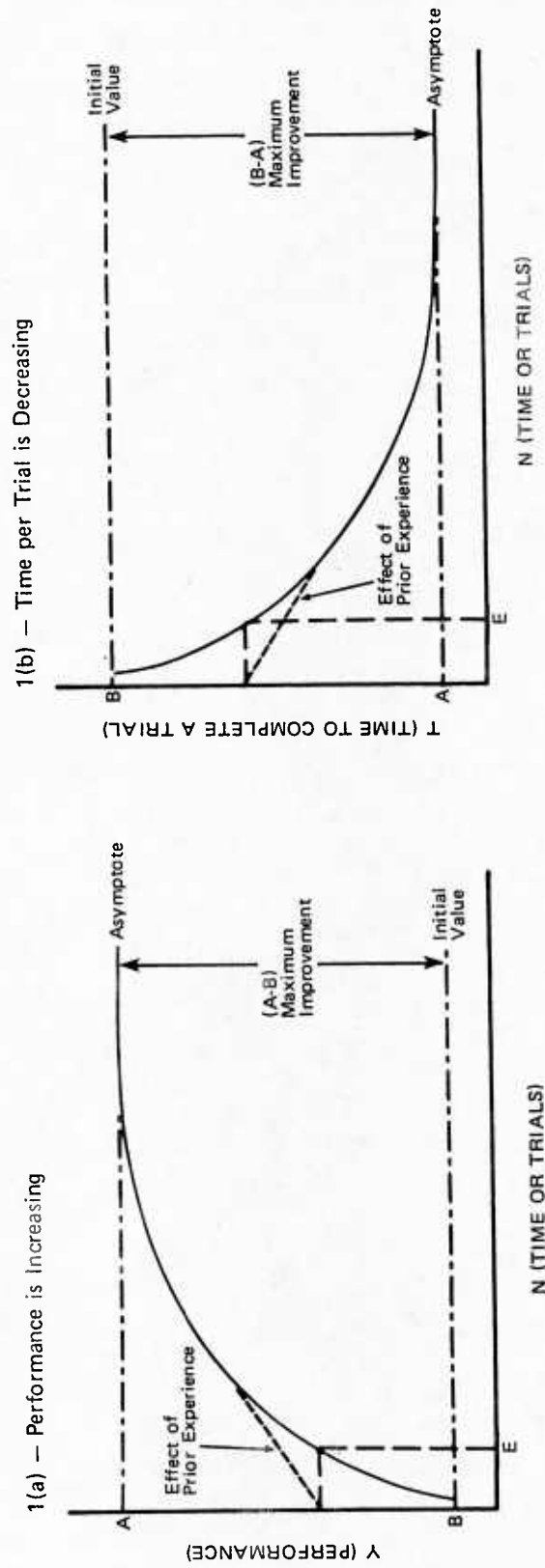


Figure 1. — PARAMETERS OF TYPICAL ACQUISITION CURVES

The functions below represent the "big three" ordinarily proposed to describe the pattern of learning data over time, along with the logistic equation. Presentation of the first three generally follows (with some notational changes) that of Newell & Rosenbloom (1981); the logistic follows Spears (1983; 1985), but is recast into our notational schema. Other curves to be discussed later are variants of one or another of these forms. The first three functions are cast in terms of T (time to respond), which produces a decreasing curve and for which the A parameter (asymptote) is the lowest possible score; these equations also hold for increasing functions (signs will change for some parameters). Some variants of these functions presented below are given in their "most usual" form which involves an "increasing performance" shape. (As the various curve families are discussed below, the reader may wish to preview a later section on comparative analysis of functions. Figure 3 in that section compares the fits of four different curves to a single data set.)

Generalized Power Function

In its most complete version,

$$T = A + B (N + E)^{-R} \quad \text{(Power function)} \quad [1]$$

The power function is sometimes encountered in a simpler form with the presumption that A and E are zero.

$$T = B N^{-R} \quad \text{(Power function)} \quad [1a]$$

Curves that follow the power law show the characteristic that the change (improvement) in performance between two trials decreases systematically as the number of trials increases. The rate variable, R, is a measure of how rapidly improvement drops off as a function of practice and indicates the degree of curvature of the learning curve, the rapidity with which

asymptote (if any) is reached. The decay (or increase) under the power law is such that if T (or Y) changes by a given factor (e.g., 2) over n trials, it will require another $n(n-1)$ trials for T to change by that factor again (Newell & Rosenbloom, 1981). For power law curves (and most of the curves posited to describe learning), the amount of learning on each trial is a constant proportion of what remains to be learned, producing a negatively accelerated curve.

Power function fits have been proposed for cumulative response curves, as well as trial by trial performance. Stevens and Savin (1962) suggest that the power function is the most appropriate descriptor in continuous response tasks such as tracking, for which division into trials is arbitrary and performance is momentary, and in experiments in which the cumulative total responses across time are the principal variable of interest (as in Skinnerian curves). They replace T (or Y) in the conventional power equation with the cumulative or integrated performance variable P . Newell and Rosenbloom (1981) also comment on the appropriateness of application to cumulative curves, and show that the cumulative power function is a conventional power law equation.

The power function has had an extensive application in industry in the form of manufacturing progress functions or industrial learning curves, which describe the increase in productivity expected to occur in a production process as a result of accumulating experience with the process (Conway & Schultz, 1959; Nanda & Adler, 1977). The earliest use was reported by Wright (1936) to estimate cost to produce aircraft as a function of cumulative production. He used a simple power law function [1a] with B as first unit cost and Y as cost per unit after N units produced. As production experience was gained during World War II, Wright's "cumulative average learning curve" was found to be inadequate for much empirical data, in that it did not provide for the benefits of experience

gained by building similar aircraft in earlier production runs. The "Stanford Curve," a modified power law, replaced the Wright equation (Nadler & Smith, 1963; Nanda, 1977). The Stanford Curve below [1b] is equivalent to [1] without the asymptote parameter, and is probably the most widely used equation for production estimation. Note that manufacturing progress functions are intended to apply to the production system as a unit, and only secondarily to the performance of individual operators or other system components.

$$Y = B (N + E)^R \quad \text{(Power function)} \quad [1b]$$

Generalized Exponential

$$T = A + B e^{-RN} \quad \text{(Exponential)} \quad [2]$$

where e is the natural logarithm. The exponential, as later variants will show, can be recast into a variety of forms. As presented by Newell & Rosenbloom (1981), the exponential is substantially different from the power law. In particular, the exponential decreases or increases much more rapidly, since the amount learned on each trial does not decrease as a function of N. In general, if T decreases by a specific factor in n trials, it will take only n more trials to decrease by that factor again (as opposed to n(n-1) trials for the power law). The exponential in this form produces a curve of constant acceleration, producing a much "steeper" curve than the power law.

Other variants of the exponential are commonly encountered. The exponential growth curve or negative exponential takes (usually) a negatively accelerated form similar in shape to the power function, with the increment to Y systematically decreasing across trials. As described below, the function increases across time. In our notation,

$$Y = A [1 - e^{-R(N+E)}] \quad \text{(Exponential)} \quad [2a]$$

where R has the conventional meaning. Note in this version that the fit is based only on the asymptote, without use of initial value. The portion in brackets reflects the percentage of the asymptotic value attained on each trial. This equation is also at times used without the prior experience parameter.

$$Y = A [1 - e^{-RN}] \quad (\text{Exponential}) \quad [2b]$$

A varying notation of the same curve employs a different interpretation of the rate variable. In [2c] and [2d] below, the parameter t is a rate constant related reciprocally to R, such that large values of t are associated with slower increases in Y. These curves are the same structure, with notational changes, as those reported by Mazur & Hastie (1978).

$$Y = A [1 - e^{-(N+E)/t}] \quad (\text{Exponential}) \quad [2c]$$

or

$$Y = A [1 - e^{-N/t}] \quad (\text{Exponential}) \quad [2d]$$

As with the power function, use of the negative exponential in various forms is common in the industrial engineering literature. A version called the "time constant" model (Towill, 1976) involves both A and B parameters (in our notation):

$$Y = B + (A - B) (1 - e^{-N/t}) \quad (\text{Exponential}) \quad [2e]$$

where t is the time constant with the same meaning as that defined above. The time constant model is most typically applied to the performance of individual operators rather than to the total production system as in the manufacturing progress function. Note that the quantity (A - B) is the difference between initial performance and asymptote, and represents the maximum increase in performance due to learning. A variation on

the time constant model is given by Johnson (1980), although attributed to other authors,

$$Y = B + (A - B) [1 - e^{-R(N-1)}] \quad (\text{Exponential}) \quad [2f]$$

where R has the conventional meaning. The rationale for the (N-1) term vs. N is not given.

A number of common curves from the learning literature are cast in exponential form. As an example, Hull's (1943) well-known equation for habit strength is identical in form to [2b]. In Hull's notation,

$$H = m (1 - e^{-iN}) \quad (\text{Exponential}) \quad [2g]$$

where H is habit strength (analogous to performance), N is reinforced repetitions, m is asymptote, and i is a rate variable. This equation, with a constant reflecting initial performance, has been frequently used to describe learning performance (see Digman, 1959).

Hyperbolic

$$T = A + B / (N + E) \quad (\text{Hyperbolic}) \quad [3]$$

or, cast as a special case of the power function,

$$T = A + B (N + E)^{-1} \quad (\text{Hyperbolic}) \quad [3a]$$

In its simplest version, the hyperbolic equation involves only two fitted parameters, and is commonly used in that form.

$$T = A + B / N \quad (\text{Hyperbolic}) \quad [3b]$$

For all hyperbolic fits, positive B yields a decreasing function, negative B an increasing one. Note that the

hyperbolic equation involves an implied R (rate variable) of -1, and thus requires the fitting of one less parameter in any form than its parent equation, the power function. With that rate, it produces a steadily increasing (or decreasing) curve of varying instantaneous slope which changes between successive trials in accordance with the ratio $n/(n+1)$.

The hyperbola has likewise been applied in the learning literature. Mazur & Hastie (1978) discuss its use both with and without the prior learning parameter. They proposed a variant which involves the asymptote rather than initial performance as a departure but still takes hyperbolic form. In our notation,

$$Y = A [(N + E)/(N + E + t)] \quad (\text{Hyperbolic}) \quad [3c]$$

where t has a meaning similar to that for the exponential (version [2c]). It indicates the rate of approach of the function to the asymptote A , with a large value indicating a slow growth.

Other curves of hyperbolic form have been suggested. In one of the earliest mathematical descriptions of learning curves, Thurstone (1919) fit cumulative errors (U) with the following function (original notation):

$$U = R / [a + (akR / m^{1/2})] \quad (\text{Hyperbolic}) \quad [3d]$$

where a and m are constants, R is number of trials (our N) and k is the probability that a given error is the last.

Logistic

$$Y = A / [1 + (B - A) e^{-kN}] \quad (\text{Logistic}) \quad [4]$$

where k is implicitly a function of R such that $k = R / [Y (A - Y)]$. Note that k in this context, although fit as a constant,

is not the same constant rate measure defined for the other curves. It is, as is rate for the power law, a measure that varies across trials, reflecting a rate that changes in accordance with amount already learned, but in a different way than in other curves presented. R has the meaning of a proportionality constant, indicating the percentage of learning yet to be accomplished acquired on each trial; k contains both that constant and an accelerative component based on amount already learned. The greater the learning thus far, the faster learning occurs. Previous learning has a "catalytic" effect on subsequent learning; for this reason, the logistic curve is sometimes called the "autocatalytic equation."

The logistic equation produces generally S-shaped or sigmoid curves, which rise slowly in early trials, accelerate rapidly in the mid-portion, and level off, becoming progressively flatter as Y approaches A . A "complete" logistic curve will tend to have two inflection points rather than the one characteristic of power or exponential equations. Spears (1983; 1985) places considerable emphasis on inflection points of both exponential and logistic curves as important parameters in evaluating training progress. In the logistic equation [4], it should also be noted that B (initial performance) could contain a component due to prior experience on the task in question, and the N parameter in such a case would more properly be replaced by $(N+E)$, with E measured in trial equivalents of prior experience. If there is sufficient prior experience, the obtained curve will likely not show the early slow rise, will not have the first inflection point, and will appear more similar to the power and exponential curves.

Linearity Transforms

We have referred casually in previous sections to the regularity of "log/log linear" descriptions of acquisition data. This concept of fitting the logarithmic transforms of X

and Y variables, while more commonly associated with power functions, is applicable (at least theoretically) to all the curves described above. There is a tendency in all of the above curves, when the abscissa (time) variable and/or the ordinate (performance) variable are in logarithmic form, for the fits of the transformed data to be linear in shape. If transforms are performed on one variable alone (usually but not always the performance variable), the resulting shape is said to be "semi-logarithmic" or semi-log. If transforms are performed on both axes, the fit is logarithmic or log/log linear. All the basic functions above can be cast into a logarithmic framework, although some make greater interpretive sense than others in that form. The general equations are (in natural logarithms):

$$\text{Log Power: } \log(T - A) = \log(B) - R \log(N + E) \quad [5]$$

$$\text{Log Exponential: } \log(T - A) = \log(B) - RN \quad [6]$$

or

$$\text{Log Negative Exponential: } \log(1 - Y/A) = -RN \quad [6a]$$

$$\text{Log Hyperbolic: } \log(T - A) = \log(B) - \log(N + E) \quad [7]$$

$$\text{Log Logistic: } \log(A/Y - 1) = \log(A - B) - kN \quad [8]$$

Note that the exponentials and the logistic are actually semi-log in form. The performance variable (Y) is in log terms, while the time variable (N) is not. They would thus graph as linear when the ordinate is in log performance and the abscissa in conventional time units. The power function is log/log linear with the slope of the line as the rate R, as is the hyperbolic, with slope = ± 1 .

It is important to note that the functions above, while they are generally curvilinear (not straight lines), are not nonlinear in the mathematical sense. Exponents for coefficients in all terms are to power 0 or 1. For example, B as a

coefficient, regardless of the term which it multiplies, produces linear equations; B^2 or higher would produce a nonlinear system. All the functions can be cast into a linear framework of the form $y = a + bX$ (a is intercept, b is slope), where X takes the following values:

Power: $X = (N + E)^{-R}$

Exponential: $X = e^{-RN}$

Negative Exponential: $X = A e^{-RN}$

Hyperbolic: $X = 1 / (N + E)$

Logistic: $X = e^{-kN}$, where $y = (A/Y)$, $b = (B-A)$ and the intercept $a = 1$.

There is, so far as we aware, no inherent advantage to the use of equations in linear or logarithmic form other than that the display of fitted curves in that format makes apparent the steady drop in performance improvement per trial as additional practice is acquired.

Positive vs. Negative Acceleration

Acquisition functions in the families above yield one of the three general shapes shown in Figure 2 -- negatively accelerated, positively accelerated, and one which has both positively and negatively accelerated components. Most common in learning curves (except for the early stages of the logistic) is the negatively accelerated curve. A negatively accelerated curve is concave downward for decreasing functions and concave upward for increasing ones. It is possible, however, to obtain acquisition curves that are positively accelerated. For such curves the amount of change in performance on each trial increases with practice throughout the range of performance,

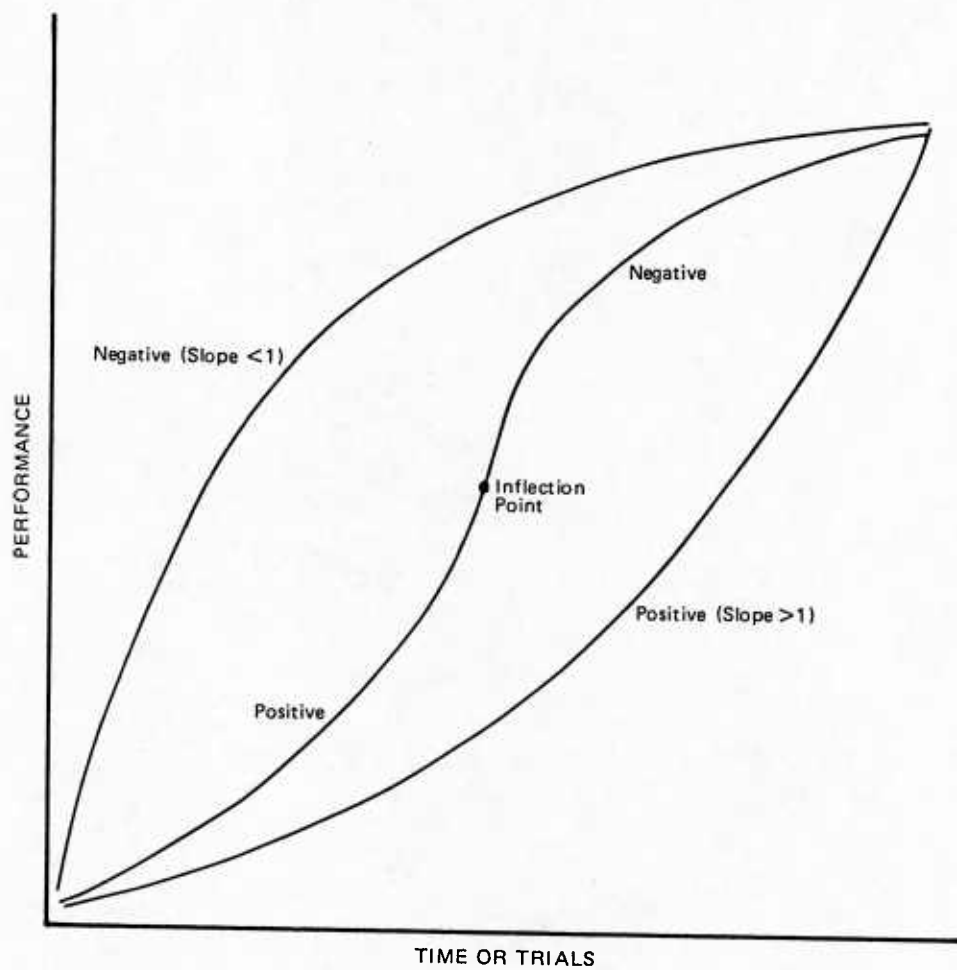


Figure 2. – CURVES OF POSITIVE, NEGATIVE AND CHANGING ACCELERATION

similar in shape to the portion of the logistic curve prior to the first inflection point. Performance estimates from positively accelerated curves rise without theoretical limit.

The functions in the previous section all produce curves with positive acceleration when the absolute value of the rate variable (the "slope" of the curve) exceeds 1.0. All previous discussions of curves are based on the presumption that the absolute value of R is in the range 0 - 1. This is primarily because the converse, a learning function which becomes progressively steeper without limit as a function of practice, is not typically encountered and makes limited sense interpretively. The value of R less than 1 causes (with the exception of the generalized exponential) the decrease in increment over time. A slope exceeding 1 implies that each increment will be larger than the one before. This is equivalent, as previously noted, to building in a cumulative way on prior learning as in the logistic. The logistic curve can in fact be viewed as one with initial positive acceleration, transitioning to negative acceleration as learning approaches some maximum possible value. Since it involves an explicit asymptote parameter, the logistic does not increase without limit.

COMPARATIVE ANALYSIS OF ACQUISITION FUNCTIONS

Several different general families of curves - power (including the hyperbolic), exponential and logistic - have been proposed as the "basic" or common shape function for acquisition, each with some empirical basis. We summarize next the comparative evidence on "which curve fits best," and why that might be so.

There is, throughout the literature, a general agreement that acquisition of a skill typically shows a pattern in which there is a decreasing amount of improvement per trial with

increasing trials of practice. As Glover (1966) suggests, "Most experimenters incline to the view that learning of a complex skill may be represented by a curve having negative acceleration with, perhaps, one or more plateaus at intermediate points." (p. 43). We will note in later sections that at least some authorities would not concede the existence (or at least the importance) of plateaus (Keller, 1958; Newell & Rosenbloom, 1981), and we will identify some conditions (particularly task difficulty) under which curve shape can take a positively accelerated form. In discussions which follow, we will for all practical purposes ignore curves of exclusively positive acceleration; they are rarely found in "practical" learning situations and reflect a poor choice of task characteristics more than an important generality about acquisition.

Curves of negative acceleration can occur under a variety of different assumptions about the underlying processes of skill acquisition and can be described by several different families of mathematical functions. They are by no means the only shapes encountered, but they recur in the literature with a regularity that cannot reasonably be ignored in either understanding or predicting the course of learning new skills. As we noted previously, the only other shape observed with any frequency is the logistic family. Spears (1983, 1985) reports on several learning studies in which data are fit almost perfectly by the logistic equation, and not well by other methods. Data in these analyses all involve ratings of performance by instructors as opposed to actual measurements, and it is likely that there are properties of rating data distributions that predispose toward the logistic form. Schneider (1984) suggests that such a finding is characteristic of skill assessed in a rating format. We believe it likely that instructors (particularly military ones) apply a moving expectation by which performance is judged; in very early trials, there is a tendency to underrate progress to reinforce the importance of perfect error-free performance. As practice increases, performance begins to be slightly

overrated to motivate continued effort towards sharpening skills collectively. As performance approaches the desired level, high ratings (mostly very high marks) are given to all trainees, and attention shifts to the refining of detailed deficiencies. This mechanism would produce the logistic curve for ratings data. It should also be recalled that the part of the logistic after the inflection point is a negatively accelerated curve, and, if no positively accelerated portion is present, the logistic is usually indistinguishable from the other curves described. We will thus in our comparisons focus on the evidence in support of the negative exponential and power law forms, and attempt to contrast evidence on these more frequently advocated functions.

The "Universality" of Learning Functions

It is important, in the examination of what curve fits what data and in the search for generalizable mathematical descriptions, to distinguish between abstracting regularities which may then generate further analyses and using these regularities to imply some general "laws of learning." There is virtually no support among modern theorists for the concept of universal laws to explain skill acquisition; there was considerable dissent among investigators even when such efforts to derive generalized underpinnings were in vogue in an earlier period. Shepard and Lewis (1950) concluded that "...there is no single 'generalized' learning function, no 'true' curve of learning and performance. Presumably, the different curves presented by different investigators...have each reflected the course of learning as determined by many different factors, some of them inseparable from the learning process itself." In the industrial context, Conway and Schultz (1959) maintain that "...there is no such thing as a fundamental law of progressNo particular slope is universal, and probably there is not even a common model."

It should be recognized that these writers were reacting to an environment in which learning tasks of extraordinarily

diverse characteristics were being forced to fit a single explanation of their acquisition form without regard to learning content or the situations under which learning/training was being conducted. Within the behavioral community, the Hullian and Thurstoneian traditions were still prevalent. For industrial applications, not only form, but intercepts, rates and asymptotes had been posited as universal constants across a wide variety of task content. Nadler and Smith (1977) reinforced the futility of substituting generalities for a detailed study of the course of learning for the individual task. They showed that progress functions (particularly the rate at which full productivity was reached) differed across products, across facilities and across companies. Further, rates varied for the identical tasks at varying facilities within the same company, and for the same task at different companies. Data reported by Nadler and Smith, while it negated the use of fixed learning parameters, supported the presence of a common shape for pooled learning performance (in their case, the power function).

An Information Processing Viewpoint

In one of the current views, the consistencies encountered in learning data across diverse material are considered not so much as universal laws of behavior but as evidence that there may be basic ways of storing, processing and retrieving information which represent fundamental elementary "building blocks" in human learning and which function similarly across a wide variety of types of information. In this approach, both regularities in learning patterns and departures from regularity are considered as clues from which the nature of these processes might be further understood. While this is superficially the same goal as earlier attempts to derive general laws of behavior, there are substantive differences in willingness not only to deal with the undeniable presence of individual differences but to develop explanations for learning that are

consistent and robust to differences among individuals and task conditions.

Yen (1978) suggested that models using information processing (IP) skills as the basic components for learning offered a means of describing the individual strategies observed in the acquisition of complex tasks. Variations in individual processing parameters would account for differences in individual curves and reveal areas in which instruction in specific components or special strategies might be profitable. Similar positions with respect to IP skills have been taken by Anderson (1982) and Neves and Anderson (1981), along with many others. This approach is equated with the predominant subject matter of cognitive psychology by Gagne and Dick (1983), and is described in detail in the instructional context.

We noted previously, and discuss in detail later, the differences that can occur between group curves and the individual curves of which they are composed. The IP explanation of learning in terms of basic processes for dealing with information is consistent with both the regularities in shape encountered in group curves and the differences in patterns and strategies observed in individuals in their progress toward task mastery. Processes represent components of a multicomponent task. Individual trainees will vary across trials in their momentary emphasis on one or the other of the basic processes in the course of integrating the components. Their individual performance patterns will thus vary in accordance with both the level of ability each trainee possesses on each process and the unique pattern by which those processes are combined in progress toward task goals. This mechanism accounts for differences in individual curves.

At the same time, the amalgamation of individual performances based on a common underlying set of processes is analogous to the summation of composite variables in

psychometric terms. This end result of summed processes will produce a set of individual scores which summate on any given trial to a value behaving in a lawful and predictable way with respect to group progress, thus accounting for the consistent regularities in group learning curves. Newell and Rosenbloom (1981), in positing the "ubiquity" of the power law, suggest that this very ubiquity across varieties of material requires some form of multiple process explanation. They provide as an example a theoretical development of the "chunking" of material in memory that results in a power law explanation of learning.

We have noted previously that two major functions for curve shape have received extensive advocacy as the "basic" form of acquisition progress. The negative exponential and the power function involve slightly different assumptions about the way in which underlying processes are combined to produce learning. In the next sections we will present the evidence supporting each approach.

The Power Law

Newell and Rosenbloom (1981) performed by far the most extensive analysis of the form of acquisition curves yet conducted. They reviewed data from dozens of studies, comparing generalized power function, generalized exponential and hyperbolic fits to the reported curves. Studies included learning of both relatively short and long duration, and covered industrial tasks, tasks with predominantly motor requirements and tasks with extensive cognitive components. They found the power function to be highly descriptive of virtually all the data examined, and clearly superior to the generalized exponential. The hyperbolic (a special power function) performed nearly as well in most cases as the general power function (recall that the hyperbolic has a presumed rate value of 1.0 and thus requires fitting of one less parameter). Newell and Rosenbloom conclude that "...There exists a ubiquitous law of practice. It appears to follow a power law."

Other analyses of learning literature have reached much the same conclusion. Anderson (1982), following Newell and Rosenbloom, demonstrates the application of the power law to cognitive learning performance. Stevens and Savin (1962), examining 12 sets of cumulative learning data from predominantly motor experiments, concluded that "...the power function is a rather general finding," and that "...the only appreciable deviation from the fit of the lines is at the beginning of curves. Departures are of the kind expected if the learner started with some degree of skill." Mazur and Hastie (1978), found the hyperbolic (with a prior experience parameter) to provide better fits than the exponential in 47 of 56 perceptual-motor learning data sets, and in 21 of 23 sets of verbal learning data (some of these were also used by Stevens and Savin). Fitts and Posner (1967) suggest the power function as the most appropriate for use in describing motor learning. Gentner (1983) found essentially perfect fits of typewriting skill acquisition to the power function (in its log/log form).

In the industrial community, the power law has been virtually synonymous with variants of the manufacturing progress function (Conway & Schultz, 1959; Nadler & Smith, 1963; Nanda & Adler, 1977; Wright, 1936). Hundreds of manufacturing learning curves have been reported as appropriately represented by the power law, and in its present form (involving an experience parameter) it is still widely used in industrial applications.

Newell and Rosenbloom present several conclusions about acquisition curves from their theoretical and empirical analyses.

a. Empirical curves do not fit the exponential or logistic form. The tails (late practice in their decreasing curves) are slower than exponential fits would provide, and the adjusting of asymptote parameters cannot totally compensate. There is no

trace of an S-shape in any of the reported curves, and the logistic (sigmoid) is not required to obtain satisfactory fits.

b. The data fit the generalized power function (including hyperbolic) with little shape variance remaining to consider alternative shapes.

c. The data do not fit the simple power law. Asymptote and experience parameters are required.

d. It cannot be determined if the hyperbolic fit is the common form or if the general power law is required (recall that the hyperbolic uses fewer parameters).

e. The power law fits all types of cognitive data as well as perceptual-motor data.

Exponential Equations

Not all writers who have analyzed learning progress accept the "ubiquity" of the power law. The exponential growth equation or "negative exponential" produces curves which show the same negative acceleration as the power function and are visually much the same in appearance. There is nearly as large a literature supporting the negative exponential as the common form of skill growth as that previously cited for the power function.

In the behavioral learning literature, all the work of Hull (1943, 1952) used exponential forms for the growth of habit strength, as did that of Wickelgren (1974) for habit strength decay. Virtually perfect fits to the negative exponential have been reported for a wide variety of learning data. Digman (1959), and Noble, Salazar, Skelley and Wilkerson (1979) give examples of application to conventional perceptual-motor tasks. In a slightly different context, Knerr and Sticha

(1985) and Rowland (1985) successfully used exponentially-based models to describe respectively the learning of procedural tasks for armor and the increased detection of tanks occurring with practice.

When data are fit well by both exponential and power functions, it is frequently difficult to conclusively choose the "best" fit. Much of the data reported in support of the power law also support the exponential as a satisfactory descriptive function. Of the 56 perceptual-motor curves fit by Mazur and Hastie (1978), 47 fit the power (hyperbolic) function "better." Of those 47, however, the difference between hyperbolic and exponential fits was less than 0.25% in explained variance for 25 curves (many of those differed only by rounding error, less than 0.01 percent), and less than 0.50% for 36 of the 47. In only 8 of the 56 cases was there a difference in explained variance greater than 1%, and half of those favored the exponential. The superiority of the hyperbolic function for the Mazur and Hastie data was thus far from conclusive. Both functions required the fitting of three parameters. As Mazur and Hastie point out, increasing the number of parameters estimated provides progressively better fits but makes relative superiority of functions more difficult to determine.

Although the power function (usually in its log/log version) has received the bulk of attention in the generation of manufacturing progress function, it has not received total acceptance, and a variety of alternatives based on the negative exponential have been proposed.

Towill (1976) advocates the "time constant model" as the common form for the majority of industrial learning curves. The time constant model, as given in a previous section, is an increasing exponential involving both asymptote and prior experience parameters. Towill describes a variety of industrial production and worker output curves that take the time constant

form. He reports work by Hackett (1974), who compared time constant fits to 12 other potential curve forms (not specified in detail, but presumably including the power function) for 88 industrial data sets. Hackett found the time constant model to fit all 88 sets as well as or better than any of the other forms.

Other writers from the engineering community have questioned the appropriateness of the power law for all applications. Johnson (1980) proposes a variant of the time constant for determining the most cost-effective termination point for training. Levy (1965) views both training and practice on tasks as a form of adaptation to the task environment. He presents a variant of the exponential which in theory should be more pertinent than the power form for tasks involving steeper learning curves and in which performance is more readily modifiable by training or other manipulation of the learning environment. Pegels (1969) expands on Levy's concept, and argues that the exponential forms are applicable to a wider range of industrial tasks than the power law. Extension of Pegels' argument would imply that the apparent ubiquity of the power law in the industrial community may be due in part to an overrepresentation of assembly, machine operation and other heavily manual tasks in the data which support the power law, and an under-representation of tasks involving significant planning and integration components. Although data in both Newell and Rosenbloom (1981) and Anderson (1982) suggest that the power law is applicable to tasks with high cognitive content; virtually no data are available comparing the relative effectiveness of negative exponential and power fits on the same cognitively-oriented data sets.

Contrasting the Functions

Inconclusiveness of the literature -- The only consensual statement that is justified by all the evidence is the quote from Glover (1966) which initiated our discussion of comparative

analysis -- virtually all the data support the contention that learning curves in general take a negatively accelerated form. Whether that form is universally the power law or the exponential is by no means conclusively shown by the data cited above. Although there is a somewhat greater volume of supporting studies for the power law family, there are also hundreds of studies and equally sound logic in favor of the exponential. Some of this difficulty lies in the perceived requirement to make an either-or choice between the two alternatives. It is probable that one is better under some (as yet unknown) learning conditions, the other under different (equally unknown) conditions. The literature is diverse and essentially uncontrolled with respect to the many factors that are known to have an effect on performance and on curve shape. There likely are ways of combining task characteristics, learning environment conditions, method of instruction and so forth into clusters for which one or the other of the functional forms might be more clearly appropriate, but such combinations are not apparent from the examinations conducted in the present study.

Some additional data -- It is uncomfortable from a scientific viewpoint to contend with data that are almost equally strong in support of two apparently incompatible alternatives. (It matters somewhat less from a practical viewpoint, as we will point out later). In an attempt to derive our own clues as to where some resolution of the incompatibility might lie, we applied several of the common functional forms to a small number of learning data sets that were readily and conveniently available from our own work and that of our colleagues, and from the open literature where sufficient data were reported. These are representative of very nearly the complete range of task types likely to be encountered in learning studies. We make no pretense of random selection, but are convinced that no conscious bias was exercised in choosing

data for analysis other than that of obtaining the widest possible variety of tasks.

The functions examined were the power function, the hyperbolic, the negative exponential, and an additional equation, $Y = a + b [\log(N)]$, which is of linear form with a log transform of trials. The first two are both log/log linear, the latter two are semi-log linear, with the exponential being logarithmic in Y and the latter equation logarithmic in N. All equations used involve the estimation of two parameters.

The data sets examined and important aspects of the analysis outcomes are outlined in Table 1. The fits of the functional forms are not unequivocally in support of any one equation. They support in part the "ubiquity" of the power law and the hyperbolic. The two functions taken together provide the best fit or nearly so to 11 of the 15 data sets, far more often than the exponential (we will discuss the log (N) form later).

It should be noted that variance explained by functions for the rather small data sets in Table 1 should be very high. Equations which explain less than 96% to 97% of the variance for the limited degrees of freedom involved will allow for substantial discrepancies between actual and estimated points. Except for those distributions with significant plateaus, a "good" fit should explain over 98% of the variance. While a reasonable level of error is tolerable for prediction of performance, different decisions about the "best" functional form have dramatically different implications about the nature of basic processes underlying learning. Departures from fit in early and late portions of a curve that are unimportant for estimation may be critical in differentiating between alternative explanations of, for example, whether and at what rate performance approaches asymptote.

TABLE 1

Goodness-of-Fit of Several Learning
Functions to Selected Data Sets

Reference	Dependent Variable (Y)	R-Squared fit by			
		Power	Exp.	Hyp.	Log(N)
Ehrlich (1943) (12 pts) ^a	Fencing lunge accuracy	.980	.798	.916	<u>.991</u> ^b
	Fencing lunge speed	<u>.996</u>	.889	.900	<u>.995</u>
Bond (1985) (8 pts)	Keystroke speed (Kanji)	<u>.986</u>	.849	.842	<u>.986</u>
Jones & Kuntz, (1985)(19 pts)	Mirror drawing (tracing speed)	.949	.940	.823	<u>.977</u>
Digman (1959) (13 pts)	Time on target Pursuit rotor	.965	.873	.923	<u>.995</u>
Kieras & Bovair (1985) (10 pts)	Reading time	.736	.890	<u>.962</u>	.927
Kennedy, Wilkes, Lane & Homick (1985) (4 pts)	Paper-pencil motor test (No. Correct)	.938	.817	<u>.995</u>	.948
	Paper-pencil cognitive test (No. correct)	.903	.764	<u>.985</u>	.912
Lintern, Thomley, Nelson & Roscoe (1984) Simulator bomb delivery (10 pts)	RMS Altitude Error (Dive)	.935	.789	<u>.964</u>	.908
	RMS Altitude Error (Cone)	<u>.974</u>	.845	.963	.955
	Bomb miss Distance	.871	.850	.867	<u>.900</u>
<u>Sheppard (1985)</u> Simulator carrier landing (mean glideslope error) (6 pts)	Middle segment (whole trng)	.915	.761	<u>.962</u>	.831
	Middle segment (part trng)	.942	.865	<u>.953</u>	<u>.957</u>
	Close segment (whole trng)	.919	.768	<u>.948</u>	.802
	Close segment (part trng)	.941	.860	<u>.953</u>	<u>.947</u>

^aNumber of pairs of observations on which curves are based^b Highest R² and those within .01 are underlined

To illustrate the degree of correspondence to data represented by various values of R^2 , data from Jones and Kuntz (1985) are shown in Figure 3, along with the the plotted fits of the power, hyperbolic, exponential and log (trials) curves. The reaching of the asymptote too rapidly by the hyperbolic and the inability of the exponential to fit the early trials is characteristic of these functions on many data sets. Note also the tendencies for the data points themselves to "bounce" up and down slightly. These irregularities, while typical of empirical curves, limit the goodness of fit for any equation, and the 0.98 fit of the log (N) function seen in Figure 3 is impressively high.

The relatively poor showing of the exponential form is surprising considering the magnitude of literature supporting the exponential or some variant as the common form of learning curve shape. The exponential approaches a satisfactory fit for only one of the 15 curves, and is relatively ineffective for the majority. While it provides "significant" fits to all of the data sets, it is clearly less appropriate than the other forms for representation of the curves. To an extent, the lack of fit may be due to the version of the exponential equation used for analyses; it involves only two parameters, fewer than the equations for forms typically reported in the literature. The time constant, for example, uses estimates of both initial value and asymptote as well as rate. As we have noted earlier, allowing additional parameters to vary obviously increases the degree of correspondence obtained, but at the same time reduces the discriminability between alternate equations. Had more parameters been used, it is likely that fits for all functions would have been improved, with relative superiorities remaining unchanged.

An unexpected finding is the successful representation of so many of the data sets by the "log(N)" form. It provides the best fit or nearly so to 8 of the 15 data sets. To our

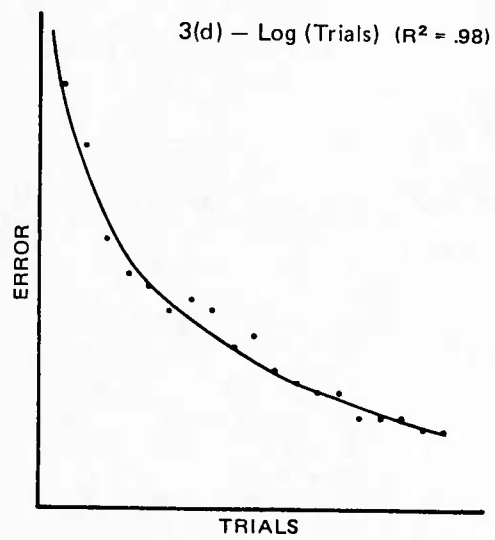
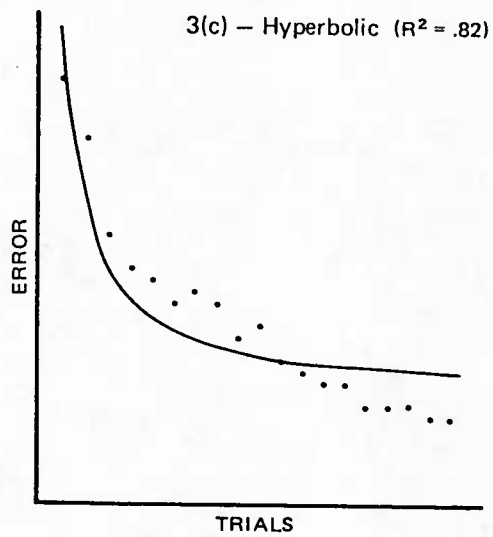
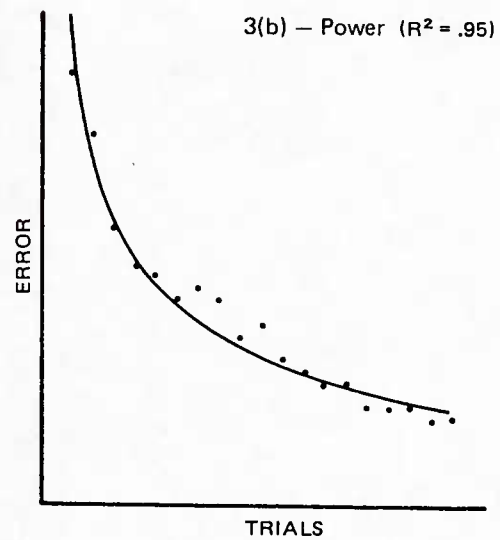
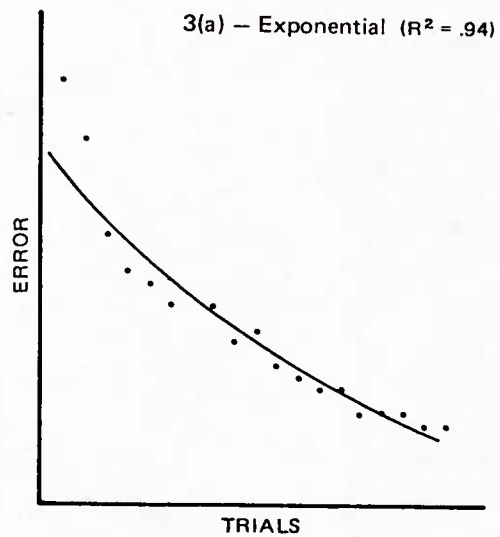


Figure 3. — EXPONENTIAL, POWER, HYPERBOLIC AND LOG (TRIALS) FITS TO STAR TRACING DATA*

*Adapted from Jones and Kuntz (1985)

knowledge, this form has not been proposed elsewhere to describe learning curves. The power function is fit in the space of $\log(N)$ and $\log(T)$, i. e., logarithmic representation of both performance and practice. The exponential involves the correlation of the log of performance (T or Y) to trials or time in original units. The $\log(N)$ form relates performance in original units to the logarithm of practice or trials.

The basis for the strong showing of that form (the "best fit" for almost half of the data sets) is not readily apparent in any conventional interpretation of the role of practice in changing performance. Fits in $\log(N)$ form indicate that the difference in performance between successive trials follows the ratio of successive times or trial numbers. Graphically, it is a "semi-log" function, in which the fit is linear when the ordinate is performance and the abscissa is log trials. Recall that the conventional exponential is also semi-log, but that the ordinate is log performance and the abscissa is number of trials. For the exponential, the difference in log performance on successive trials follows the difference between trial numbers or time.

The highly predictive fits for the $\log(\text{trials})$ equation shown in Table 1 do not offer any immediate interpretation of underlying processes. Newell and Rosenbloom (1981) and Anderson (1982) describe in detail the characteristics of learning process "primitives" which must be assumed for equations to take either the exponential or power law forms. These address exclusively the implications of assumptions about impact of processes on the performance variable alone or taken in conjunction with the practice variable (trials), and are not readily generalizable to the present findings. Given its predictive strength, it is of considerable interest that this function has not been previously brought forward as an explanation of learning.

In sum, our limited search for clarification of the "real" common form suggests that the power law and its underlying assumptions continue as the most interpretable candidate for description of performance changes with practice. The hyperbolic represents, for at least some types of data, a powerful and mathematically economical alternative to the basic power function. For the few studies we examined, the exponential is less appropriate than its broad base of support from the literature suggests. To distinguish with certainty between competing explanations of the form of learning, a far more extensive analysis than that conducted here is required, one which systematically obtains representative curves from a broad spectrum of task types with performance assessed through a variety of metrics. The definitive analysis should examine both very long and very short periods of practice, since a major differentiation among proposed forms lies in the nature of ultimate asymptotes, both in the level eventually achieved and in the rate with which asymptote is approached.

From a pragmatic viewpoint, the processes involved in generation of a learning curve may matter less than the fidelity with which changes in performance can be represented. As we will discuss in the context of curves for making decisions about training, it may be important within a course segment to have reliable estimates of trainee progress; it is much less critical to understand why progress takes some particular form.

CONDITIONS THAT AFFECT CURVE SHAPE

In the previous paragraphs we have presented and contrasted some possible functions that might be fitted to group acquisition curves. It is important to remember that these are in essence "theoretical" functions that could be considered as candidates for the "real" distribution (or mix of distributions) underlying an empirical data set. It is common to obtain performance data which are not represented satisfactorily by any of the above or by any other rationally derived interpretable function. (The author recently completed an analysis of learning data that required a quintic (fifth order) equation to achieve minimal "significance" with a sample size of 13. The intractability of such cases is not a deficiency of curve descriptions so much as a flag to look for methodological or measurement problems).

There are innumerable combinations of task characteristics and data collection conditions that can produce such an outcome. Almost certainly, the chief problem is measurement error. Beyond that consideration, however, there are properties of the learning tasks employed and performance measures used, and features of the learning environment that can cause dramatic departures of otherwise orderly curves away from traditional shapes. In the case of group curves, there are risks inherent in the summation of individual learning performances that can seriously confuse interpretation of results. Some of these conditions and their effects on curve shape are discussed below. Many of these influences have a tendency to interact in ways such that one factor can either exaggerate or diminish (and sometimes mask) the impact of another, and make isolation of individual effects (and their discussion) somewhat difficult.

Nature of the Task

Tasks for which learning data have been collected vary widely in complexity, difficulty and level of integration. Some depend mainly on motor skills, others on verbal processing, planning or a variety of other cognitive factors. Each of these underlying skill requirements could, at least potentially, develop at different rates, thus producing, for a task involving skill mixtures, an amalgam of individual growth curves reflected in a single performance curve. To the extent that successful performance involves integration of several different task components, each emphasizing a different skill, it is highly unlikely that exactly equivalent proficiency on each will be attained in equivalent time frames. In addition, the requirement for skill integration demands resources in and of itself, and periods of little or no observed growth in component skills may occur while the learner focuses on putting together the pieces of the task. This latter condition is believed to underlie the phenomenon of "plateaus" in acquisition curves. Plateaus (i. e., no noticeable improvement), their existence and their causes, are the subject of considerable theoretical and practical interest, and are dealt with separately later in this section.

In addition to its possible contribution to the occurrence of plateaus, task complexity (the presence of multiple components or kinds of skill demands in a task) has other potentially disruptive effects on finding simple mathematical representations of performance. To the extent that task components must to some extent be practiced separately and ultimately combined in some way by the trainee, the opportunity occurs for individuals to use substantively different strategies for practice and integration of the task components. This introduces two separate factors which militate against orderly findings on shape of the acquisition curve. First, it brings into play an additional skill possibly independent of the other

task components, the ability of the individual to select and evaluate learning strategies. This skill is likely to increase performance variability in a way not related to actual task proficiency, since it may be in force only during learning. Second, it enables trainees to take widely variant paths to attaining the same ultimate proficiency in the same time frame, no one of which can be legitimately represented as superior or inferior, but with each resulting in a different shape to the acquisition curve.

Isley, Spears, Prophet and Corley (1982), in a study on the training of landings in a flight simulator (an extremely "multicomponent" task) found that trainees concentrated on mastering one component (such as altitude control) at a time, allowing performance on other components to deteriorate below previously attained levels. While group performance increased slightly but steadily, individual performance was extremely unstable initially, and performance curves took the expected form (negatively accelerated) only after integration of component skills began to develop.

Ehrlich (1943) reported a similar finding in attempting to develop a learning function for speed and accuracy measures of a fencing task which involved a number of separate motor components. He found initial performance so unstable as to be unsuitable for use as a parameter in curve fitting and found it necessary to use the first three training sessions (about 30 trials) as a base for the initial performance value (B in our notation).

Finding an unstable initial value is likely far more prevalent than has been reported. Ehrlich detected the effect of early instability on curve shape because he looked for it as a part of his theoretical development. It is expected in complex tasks, is present in many other studies reviewed here, and may be responsible for many of the relatively poor fits

obtained for curves which should, on the basis of task characteristics, be reasonably well described by conventional learning functions. Differences in early trial stability may also account for the presence in the literature of some curves which are fit marginally better by the power law, others which follow the exponential law slightly more closely, and a myriad of others for which the differences between fits are too small to matter. Instability, which makes curves difficult to fit reliably by mathematical functions, almost certainly has a much heavier disruptive impact on individual curves than on group curves. Carter and Woldstad (1985) address the instability of early practice trials, and provide an excellent discussion of the role of stability in the theoretical underpinnings of analysis of individual acquisition curves. They draw heavily in their discussion on the analysis of many hundreds of learning curves summarized by Bittner et al. (1984).

The group curve produced by summation of individual ones for multicomponent tasks may be accurate with respect to initial performance, asymptote, and average rate, but it can be seriously deficient if it is used to track individual performance for possible intervention and remediation.

The nature of activities involved in successful performance of a task can exert dramatic influence on the shape of individual and group curves. A task that requires concept formation or recognizing and abstracting a rule from observational or outcome data may show no learning at all for a period of time, with an immediate leap to continued correct performance once the appropriate rule or strategy is discovered. Such "insight" or "all or none" learning produces individual curves in the form of step functions, varying among trainees only in terms of trials required until the concept is understood. As Baloff and Becker (1967) point out, summation of a series of individual insight learning curves can produce a

very satisfactory continuous negatively accelerated curve that mirrors one obtained from conventional cumulative learning.

The characteristics of the task in terms of skill requirements, particularly task complexity, can thus affect both the predictability of individual curves and the interpretation of the group acquisition rate and other parameters. The relatively simple models represented by the above functions are frequently inadequate to describe learning of very complex, multicomponent tasks. None of the commonly-used functions can deal with a curve in which there are one or more "flat spots," portions in which the change in performance from trial to trial is essentially zero; neither can any other single curve which is linear in its coefficients. It may not be reasonable to expect such fits for highly complex tasks. Mazur and Hastie (1978), for example, in their analysis of learning curves, deliberately selected simpler tasks to evaluate their theoretical position on curve shape to avoid the complications of such varying rates of change.

Task Difficulty

We noted previously the existence of positively accelerated curves and their infrequent occurrence in the learning literature. One condition which can produce positive acceleration in acquisition data is unusually high task difficulty. If successful performance requires skills that are well beyond the capabilities of the trainee, successful outcomes are few (often random) and the reinforcement value of a trial is small. It may take many trials before the beginnings of an appropriate motor program or strategy can be developed, and the amount of practice required to attain reasonable proficiency may exceed any reasonably available time.

Curves in Figure 4 are adapted from a study by Krueger (1947). They illustrate clearly the effects of task

difficulty. His task consisted of tossing rings over nails in a wall; task difficulty was varied by the distance from the wall (2, 3, 6, and 9 feet). From the 2-foot distance, the task was too easy. Curves rose quickly to perfect performance. From the 9-foot distance, the task was so difficult that learning was almost imperceptible. Across one thousand tosses, there was only a slight increase, from an average of 1 success in 20 trials to a maximum of about 1 success in 12 trials. Clearly, successful performance must occur at a reasonable rate and be "reinforced" in order for learning to occur.

For intermediate distances (3 and 6 feet), the Krueger curves show an unmistakable positive accelerated form. From 3 feet, performance rose almost linearly with slight positive acceleration for 700 tosses, changing to negative acceleration and beginning to approach asymptote at 1000 tosses. From 6 feet success increased slowly in a linear fashion for 700 tosses, rising rapidly with positive acceleration past that point through to the conclusion of the experiment. The practice period employed was quite long for a relatively uncomplicated task, but even with that extensive practice only the 2 and 3-foot groups approached satisfactory performance. Had the experiment been concluded after 500 throws, the 6 and 9-foot groups would have shown almost no learning, and the 3-foot group would have been judged as only slightly skilled. Learning was nonetheless occurring, even for the most difficult conditions. Had practice continued for 2000 or 3000 tosses, it is likely that performance improvement would have continued and, in the case of the 6-foot group, increased in rate considerably. The curve for the 6-foot group shows by the 1000th event a shape much like that of the logistic curve up to and just after the inflection point.

Krueger's data provide several powerful generalizations about task difficulty, curve shape and learning. In his study (as in most others), difficulty drives learning rate, which in

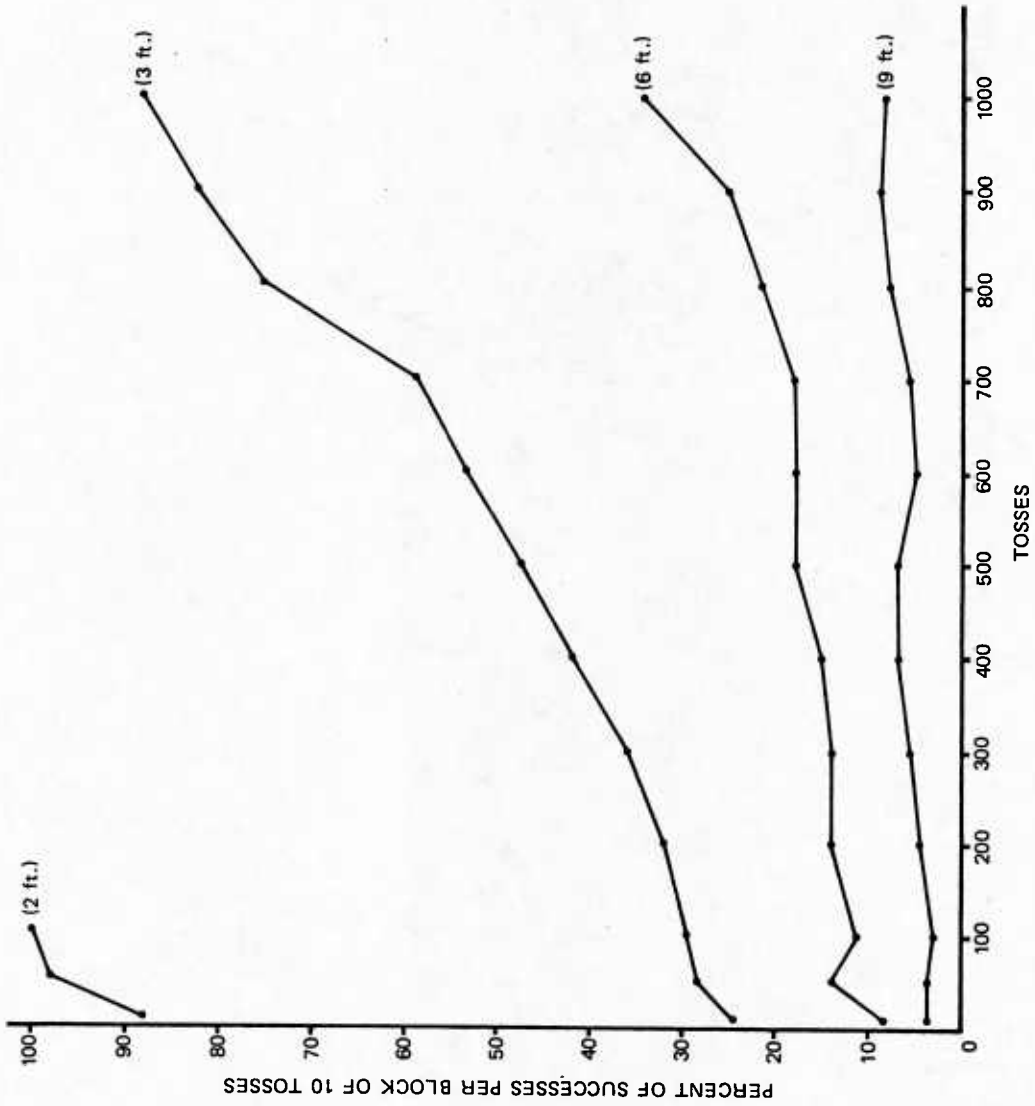


Figure 4. — THE EFFECTS OF TASK DIFFICULTY — Successes in Ring Tosses from Varying Distances*

* Adapted from Krueger (1947)

turn determines the time course required for proficiency. If tasks are too easy, or there are inherent upper limits to the performance measure, a ceiling effect can occur. This is often more subtle than in the Krueger data, where asymptote for the shortest distance was reached in about one-tenth the throws required for the next shortest. Further, some tasks as constituted may be too difficult to train in reasonable periods and simplification or segmentation may be required.

This insensitivity to practice of excessively difficult tasks, and its detection through slight positive acceleration in the curve, are pointed out also by Mazur and Hastie (1978) and by Bahrick, Fitts and Briggs (1957). The latter authors describe the "subtle artifacts" present in curves for both unusually easy and unusually difficult tasks. Finally, any arbitrary length for the acquisition period may reveal only part of the curve, and the shape of the curve during that portion may give a misleading impression. This latter finding is reinforced by numerous other studies to be discussed later, particularly Spears (1983; 1985).

Degree of Prior Learning

The extent to which trainees have experience with a given task or with some components of that task has a strong effect on the shape of the curve and the success with which a curve can be fit with one or more of the functions previously discussed. All of the commonly proposed equations involve some variant of the B parameter (trial 1 or initial performance). Inherent to those equations is the presumption that B reflects true initial capability (a "zero" skill status), since most systematically increment or decrement B by a percentage based on either B alone or the differential between B and the asymptote A. If B has a significant elevation or depression from prior experience (particularly if experience is not equivalent for all subjects),

both the wrong descriptive equation and an incorrect estimate of its goodness of fit are likely to occur.

The need to account for previous experience in curve evaluation is well documented. Shephard and Lewis (1950) make the point that few if any tasks start from a zero-learning position. Transfer is present from prior everyday common activities to virtually all new learning situations. The generalized equations above contain a parameter (E) which adjusts the number of practice trials in accordance with trial equivalents of previous experience. We have noted previously that the different initial status resulting from practice can exclude the early portion of the curve, changing the shape and, if practice is not allowed for by the additional parameter, producing systematic departures in the fit of empirical data to all of the proposed functions. Snoddy (1926), in finding virtually perfect correspondence between his data and a log/log linear curve (power law) throughout most of the range, commented on the downward bend away from the theoretical curve in early trials. The need for an "experience" parameter in Snoddy's data and in other similar curves is noted also by Fitts and Posner (1967). Many of the studies analyzed by Newell and Rosenbloom (1981), despite remarkable adherence to power functions for middle and late practice, show similar departures at the beginning of the function.

Ehrlich's (1943) study on fencing skills described above in the context of multi-component tasks is also germane to effects of prior experience. He attributes much of the instability shown in early performance to "pretraining and prior experience" on skills involving body movements similar to those required by his fencing task, and notes that learning data so affected will not necessarily follow any of the well-known curves, that "factors of pretraining negate the possibility of zero origin and exclude the lower portion of these curves." (p. 503).

Plateaus

The causes of the "flat spots" observed in some acquisition curves and, indeed, whether such outcomes are a meaningful component of curves that should be addressed, have been the subject of some controversy in the learning literature. A number of group curves demonstrate, usually in the "middle" of acquisition, a temporary "leveling off" in which performance fails to increase (but rarely decreases), followed by the resumption of performance growth in the conventional negatively accelerated manner. Plateaus found in learning to type, for example, are said to reflect periods between learning to type words rather than letters, and between typing sentences instead of words. Individual curves sometimes show the same pattern, but much more irregularly, with decreases between trials not uncommon.

Plateaus have received such attention in the literature because (a) they have substantive theoretical meaning in understanding the process of acquisition, and (b), they complicate immensely the summary description of that acquisition. Although the effects of plateaus on curve description are well understood, and there is consensus that they can be brought about by periods of time required for integration of task components, there is a lack of agreement about the extent to which they do occur or must occur.

Keller (1958) refers to the "phantom" plateau, and suggests that many of its occurrences can be attributed to methods of data collection or other aspects of the study, citing in particular the plateaus found in learning of telegraphic code by Bryan and Harter (1899), one of the earliest reported occurrences. Curves of typewriting skill learning reported by Bond (1985) from a variety of studies show no plateaus, nor do those developed by Gentner (1983), although Gentner's curves

start in the fourth week of practice. Few if any of the many curves analyzed by Newell and Rosenbloom (1981) demonstrate any periods of leveling off, although their primary aim was long term description and the analysis may not have been fine grained enough to reveal any momentary flattening. Curves analyzed by Mazur and Hastie (1978) also show no plateaus, nor do those by Stevens and Savin (1962), but the former emphasized motor learning and systematically deselected complex tasks from their analysis, and the latter, examining the cumulative power function, likely did so also.

The bulk of reported evidence supports the presence of plateaus in a variety of task situations. Kao (1937), in an extensive analysis of the plateau phenomenon, examined curves from a variety of tasks, and concluded that plateaus are "definitely" regularly present in complex (multicomponent) tasks, but rarely or never in simple tasks. Kao's results are typical of those reported by most other authors. Spears (1983), for example, concludes that plateaus "do occur and are too common to ignore."

Taylor and Smith (1956), found strong and distinct plateaus in group curves on garment assembly tasks. Their tasks were clearly multicomponent in nature, involving both motor activities and development of planning and work-pacing skills. They tracked assembly output for a number of years, finding performance improvements continuing throughout 3 to 4 year periods. In their analysis, sensitivity to plateaus was judged a critical component of evaluating performance growth; they noted that the traditional practice of averaging individual curves before each individual's initial stabilization materially distorted the shape of the group curve. Although they did not explicitly address the issue, their findings on distortion of curves by simple averaging across trials may offer a partial explanation for the failure to find clearly-defined plateaus in the other studies described above. This explanation is

supported by other evidence to be described later, particularly that of Hayes (1953) and Hayes and Pereboom (1959), which deal with the effect of selection of termination points on curve shape.

Because of the powerful effect of plateau periods in their data, Taylor and Smith also suggest that three curves or a curve with three distinct components may be required to describe the process of acquisition, one component for early practice describing a rapid initial rise, one providing for relative flatness of output for sustained periods, and one for very late periods of experience in which slight but regular improvement occurs over a very long time period (essentially an imperceptibly increasing "asymptote"). Because of the relatively long periods of flatness reported by Taylor and Smith, their data indicate the danger of too early a termination to the tracking of practice. Their plateaus were of sufficient length to be interpretable in conventional studies as asymptotic or terminal performance, although performance later rose materially beyond those levels.

Glover (1966), in a review of manufacturing progress data from industry, reported findings similar to those of Taylor and Smith. Plateaus occurred over "many long-term periods of industrial tasks," attributed by Glover to task complexity. He also noted that plateaus, in the context of industrial production, had a negative impact on output and efficiency, and that their removal represented opportunities for earlier attainment of full productivity. Evidence cited by Glover suggests that plateaus can be removed or reduced by greater attention to instruction during the learning period, using a more individualized approach to demonstrating correct methods rather than relying on on-the-job experience to gradually increase proficiency. His suggested emphasis is related to the issues of "guidance" vs. "discovery" training addressed in the educational literature (see Cormier (1984) for a summary).

Other industrial work reviewed by Nanda and Adler (1977) reinforces Glover's position of actively eliminating plateaus. They suggest that more supportive training and increased management attention to formal training can provide major increases in rate of improvement during critical learning periods and that such "intervention" (our term) can be highly cost effective. Hancock (1971), in a similar context, refers to the concept of "threshold segments" of learning (roughly equivalent to the first two stages proposed by Taylor and Smith (1956)), and suggests that time required to attain acceptable proficiency can be accelerated by a factor of 10 to 20 by increased instructional activity during these segments.

Criteria for Termination (Mastery Level)

We have thus far alluded in several sections to the impact on evaluation of a curve shape of the period over which performance is tracked, or equivalently, the point in acquisition at which collection of performance data is terminated. Given the extended periods over which performance has been shown to improve, it is likely that by far the preponderance of investigations into the form of learning curves may have been terminated too early. As Schneider (1985) and others point out, skilled performance may require many hundreds of trials to develop. Barring the presence of floor or ceiling effects induced by equipment or inherent aspects of the task, performance has been demonstrated in numerous studies to improve over thousands of trials or repetitions, sometimes with dramatic shifts in curve shape as practice or experience continues long enough. In Kreuger's (1947) most difficult conditions, for example, the eventual indications of transition to a period of rapid increase occurred only after 900 or more events. In Taylor and Smith's (1956) data, termination after a few months of experience would have given an incorrect picture of both shape and eventual asymptote.

It is obvious that training and practice must, both in experiments and in real-life training situations, be terminated at some point. It is clearly impractical to perform studies as a matter of course which take months or years to complete. It is equally obvious, however, that continuing practice beyond apparent asymptote can yield surprising reversals of interpretations about the correct point for termination. This is particularly evident when comparing alternative methods of instruction for a task.

Spears (1983; 1985) cites two studies in which materially different outcomes would have been obtained with a different choice of termination points. Martin and Waag (1978) studied the effects of motion on performance of a landing task in a simulator. The non-motion group showed early superiority, attaining their "final" level of performance rapidly and making only slight improvements thereafter. The motion group started much more slowly, with an early slow rise followed by a period of positive acceleration, producing a curve which eventually surpassed by a considerable margin the "stable" performance level of the non-motion group. The motion group was clearly dealing with a more difficult task, with the additional cues provided requiring more practice to integrate, but eventually allowing for higher ultimate performance in the simulator than the less "cue rich" environment of the non-motion group. Similar results were obtained for transfer performance in Britson and Burger (1976). In early and middle trials of night carrier landing performance, the group trained without a part-task simulator showed better performance; as practice continued, however, the simulator-trained group exceeded the non-simulator group and continued to increase its superiority throughout the balance of the study. This latter finding is common in transfer studies, and may involve the concept of "contextual interference" (Shea & Morgan, 1979) and "schema" development (Schmidt, 1975), which will be addressed in more depth later.

In the practical situations encountered in the military training/education environment, it is clearly necessary to select appropriate time periods for conducting each segment or module of training. For a large part of military training, the skills or information presented in a given segment are not directly and completely job relevant, but serve, as we discussed previously, as preparation for some following training segment or as general background for job-related duties which will require the bringing together of a number of similar skills/knowledge components. Establishing appropriate segment lengths for military courses thus presents difficulties beyond those typically encountered in other training settings.

In essence, it is desirable to find a level of performance at which instruction can be terminated with the expectation that the most effective level has been attained (given the appropriate tradeoffs), and with an acceptable likelihood of satisfactory transfer to job performance or to the next training segment. There are some significant technical issues involved with determination of "most effective level" or "optimum point" to stop a particular training segment. As later discussions will show, retention is generally improved by continued practice or training, but transfer can be materially affected through the overlearning of task-specific components which must be "unlearned" in later segments.

There are also significant economic implications associated with determining the point at which training can be safely "stopped." Continuation of training past the "optimum" point increases costs and resource use in a linear way for progressively less return on investment (ROI). Johnson (1980) and Levy (1965) present formulae based on expected curve shape which quantify the decrease in ROI as a function of continued training, and estimate the expected utility of additional gains in performance from additional training units. While use of these formulae requires estimation of quantities which are

usually difficult or impractical to obtain, they present a framework which illustrates the importance of knowing when to terminate a training segment.

We noted earlier that much of the literature on curve shape and its implications, particularly that from behavioral research, has been concerned more with understanding the processes of acquisition and looking for underlying theoretical explanations than with the use of curves and functions for determining course length. By far the majority have dealt with group performance, and addressed individual acquisition patterns only secondarily if at all. Likewise, few studies have systematically varied the length of training periods for fixed content to see what happens to performance of individuals in later task settings. While reliable descriptions of group performance curves are required to estimate best course length, they must be accompanied by some indication of individual variability around the group curve, particularly at or near the point of termination of training. It is clearly not sufficient for training in a segment to end when the group average is at a preset level if 50 per cent of the individual trainees are still outside acceptable bounds at that point in time. There are thus two related components or "parameters" of an in-course performance curve: a) overall group achievement and b) the percentage of trainees attaining some preset level of performance. While there is considerable information on the first of these components in the acquisition literature, there has been much less attention to and consequently little data on the impact of the latter, the effect of group variation.

Work on improving the effectiveness of training has traditionally looked for ways of increasing achievement or performance within a constant time frame. The objective is, in effect, to add some amount to overall group performance, and by extension, to raise the level of the poorer performers to some acceptable level. Changes to training approaches in such a

framework focus on manipulation of achievement as a variable quantity, while treating practice or exposure time as a constant factor.

Mastery training -- Over the last decade, a distinct alternative to that approach has emerged in the educational community. "Mastery level training," originally proposed by Bloom (1974), reverses the conventional relationship between achievement and time as controllable variables. In its most basic form, mastery learning establishes the degree of learning required at some fixed level, and manipulates variables such as time and conditions so that all or nearly all (a fixed percentage, usually 95%) of the trainees attain that preestablished level. Each individual is provided with instruction or practice until he achieves criterion. Mastery learning thus employs time as a variable, with achievement held constant.

The rationale underlying mastery learning is based on two quantities: Time-To-Learn (TTL) and Time-Spent-Learning (TSL) (Gettinger & White, 1979). The first, TTL, is the amount of time, practice, exposure etc. needed by an individual trainee to attain criterion performance. It is a quantity unique to the individual, and is a function of a variety of factors such as aptitude, ability to understand instruction, and motivation, as well as the quality of instruction provided. The second, TSL, is a function of the time of exposure to the material, the opportunity to learn, and the actual amount of practice for the individual (presuming a willingness to learn is present). TSL is the traditional variable of time, trials, or practice. TTL, the time required, has no direct analog in conventional studies (it is closest in meaning to rate), but is a key variable in understanding the time course of training. Advocates of mastery learning see the extent or degree of learning achieved by the individual as a function of the TSL/TTL ratio. The higher the ratio, the closer the trainee approaches to the desired level of

learning. If training is terminated when this ratio is low, retention and/or transfer of learned material is likely to be unacceptable. In this view, both TTL and TSL are required for an accurate estimate of achievement, retention or transfer or for prediction of the time course of training. This argument, by extension, would also hold that knowledge of TSL alone, as in conventional studies, would be particularly insufficient for estimating the variability around the performance curve required for training quality control.

Time-to-Learn, in the mastery learning context of variable time, is the conceptual analog of the rate variable in analysis of fixed time curves. In such a view, TTL is also the key variable in studies of Aptitude-Treatment Interaction (ATI) (Cronbach & Snow, 1976; Frederickson, 1969), since the impact of differences when ATI is present is on the time required for learning. Even so, TTL as a measure has not been widely studied (Gettinger, 1984). It is, in its current usage, a construct whose existence is virtually certain but whose quantification is difficult except in a post-hoc sense. Once the individual has attained criterion performance, it can be presumed that TTL and TSL are roughly equivalent in value. Prior to training, however, the TTL for each trainee is an unknown quantity, for which estimation methods do not at present exist. Such methods for a priori prediction of TTL would likely draw heavily on the ATI literature. TTL predictions, as well as direct measures of TTL, would be useful in determining when, and at what level of performance, training is essentially complete. It would be desirable, in mastery learning approaches, to define criteria for mastery on the basis of something beyond training (retention and/or transfer), using empirical data to determine what level of mastery leads to what level of post-training performance.

The requirement for and use of the TTL parameter, in particular the TSL/TTL ratio, has an appealing logic. Both the conventional and the mastery learning literature reflect the

tendency of trainees to learn in their own time and with their own pattern of learning. Variations among individuals in their respective ratios at a given time indicate that each is at a different point on his unique learning curve. Termination after a fixed amount of time (constant TSL) too small for all trainees to have begun the leveling-off process would leave at least some short of the "mastery" point. Recent work by Jones (1985) shows that it matters dramatically for retention where an individual is on his particular learning function when training is concluded. As Jones' data indicate, the "slope" or steepness of this function for the individual is the major determinant of how well learning will be retained during periods of no practice. While the trainee is in a period of rapid learning, estimates of both his final performance level and how soon he will attain it are unreliable. Once the individual has attained relatively constant performance and proceeds a few trials past that point, not only is retention higher than at an earlier stage, but the estimation of that retention is much more orderly. A lack of attention to the extent to which trainees are approaching their own "asymptotes" can introduce an uncontrolled error component into the analysis of group curves studied over arbitrarily fixed time periods, and may in part be responsible for the inconsistencies encountered in both the learning and the retention literature.

Mastery learning in its implementation is criterion-referenced training. There are some difficulties associated with a criterion-referenced system (termination on the basis of attained performance) that do not occur with conventional ways of tracking acquisition (termination after fixed time). Hayes and Pereboom (1959) note a number of artifacts present in curves in which performance is terminated at a fixed criterion point. They show that such a basis for stopping training capitalizes on both random and cyclical swings in performance of a trainee around his own "true" average at that point in acquisition. If his performance attains criterion as a result of this variation,

he has "learned" even though true performance may still be considerably below criterion and would require a number of additional trials to achieve a true average at criterion level. This process introduces a considerable random element, usually undetected, into the time at which criterion-based acquisition takes place. Hayes and Pereboom (1959) and Hayes (1953) present several procedures aimed at detecting and/or guarding against this outcome. Requiring several trials consecutively at or above criterion will materially increase the reliability of the "learning" trial. Hayes (1953) also suggests the use of curves which are displaced so that the final points (criterion performance) coincide for all trainees, and working "backward" to ascertain the shape of the group curve.

In brief, the point in time or the level of performance chosen to terminate a training period is an important factor in determining the form of the resulting acquisition curve. An inappropriate choice of termination criterion can misdirect findings both on the ultimate level of performance attained and on the shape of the acquisition function which best describes changes in performance with practice. This risk is particularly acute when alternative ways of training are being compared. Too early a termination can produce estimates of relative effectiveness which are seriously in error if alternatives are not comparable in cost or complexity of implementation. Too long a training period, on the other hand, while inefficient, has minimal effect on analysis of acquisition. Such a condition is likely to improve retention, but can have a negative impact on transfer through the overlearning of task-specific skills.

Distribution of Practice

The literature on effects and theoretical implications of massed vs. distributed practice is extensive, and any in-depth summarization is clearly beyond the scope of this analysis. The part which deals directly with the effects of practice

distribution on curve shape tends to be somewhat more limited but still complex to interpret because of the considerable range of lengths over which no-practice intervals can be varied. Findings suggest almost universally that the introduction of "short" periods of rest between trials tends to increase performance across trials and thus increase the rate (slope) of the learning curve as a function of number of trials. Depending on the length of the interval (and ignoring forgetting effects), these increases may or may not hold as a function of total training time. Further, it is uncertain whether differences in performance between distribution conditions necessarily represent actual differences in learning.

In studies examined by Mazur and Hastie (1978) and by Stevens and Savin (1962), there were distinct tendencies for learning curves in studies with intervals of no-practice to show higher intercepts, slopes and asymptotic values than those with continuous practice; over the same number of trials. Reynolds (1952a, 1952b) and Digman (1959) provide acquisition curves which clearly demonstrate such effects. Digman estimated that massed practice subjects in his experiment received about 60% as much practice per trial as those under distributed conditions. He also found that average performance under massed practice remained below performance under distributed practice even after more than 100 trials. This decrease in ultimate level of performance acquired in a fixed number of task repetitions appears to be a general finding; although massed practice subjects may "catch up" eventually, it may require a number of trials beyond that used in the typical experiment.

The overall effect of massed practice is to decrease the increment to performance obtained on each trial; the unifying concept seems to be that the value of a trial for increasing performance is higher when there are periods available for "rehearsal" or organization of learning obtained on a trial or within a session. These inter-trial intervals need not be

lengthy for this effect to occur. Kolers and Duchinsky (1985), using a task of reading aloud geometrically transformed text from pages, found that the time required to turn a page served to increase performance on the next page. The page, regardless of length, was the unit, and the performance increases, although they eventually manifested themselves, were not expressed in performance measures without the brief inter-page interval.

Interference and forgetting -- If the inter-trial intervals in distributed practice become "long enough," there is an opportunity for forgetting to occur, and some of what was learned will be lost between trials. This may to some extent counteract the increase obtained from practice distribution, and the effects may be difficult to separate. It is clear that forgetting must at least theoretically occur between trials, increasing with the length of interval. Its effect on curve shape would be to flatten the curve and generally depress the rate of growth of the skill. Mazur and Hastie (1978) give a straightforward analysis of the interrelationships of practice distribution and forgetting.

Interference has a complex effect on acquisition parameters. In general, interference can be viewed as negative transfer from preceding experience, and would behave quantitatively in its effects as a negative value of our prior experience parameter (E), increasing the number of trials required to achieve a given performance level, and resulting in a decrease in apparent rate of learning over a fixed number of trials. If interference is too severe (control/display reversals, for example), ultimate level of performance may be permanently depressed (Fitts & Seeger, 1953; Shephard & Lewis, 1950).

Individual Differences

The well-established tendencies for individuals to learn in different ways has traditionally been ignored in looking for the

common form of acquisition shape, or treated as an error component to be reduced by control of conditions. Although patterns for some classes of task may not show massive individual departures from group patterns, this tends not to be the case in tasks involving elements of "cognitive" learning or other strategy-based tasks or when group heterogeneity is very high (see later section on Individual vs. Group Curves). As both McKeachie (1974) and Powers (1976) suggest, individual differences are the principal difficulty in expounding general laws of learning. As work on Aptitude- Treatment Interaction (ATI) (Cronbach & Snow, 1976; Frederickson, 1969) has demonstrated, trainees at different ability levels can show acquisition curves that differ in rate, intercept and asymptote under the same method of instruction, and become more similar under different (tailored) methods of instruction. High-ability students typically do better under self-paced training, low-ability students under group-paced methods (Baldwin, Cliborn & Foskett, 1976; Taylor, Montague & Hauke, 1970). Higher ability students, regardless of method of instruction, usually have a higher initial performance, and thus tend to improve at a slower rate. Unless the intent is to use a common curve for the specific purpose of detecting these individual patterns and providing alternative or accelerated instruction, these variations in learning strategies among individuals can materially complicate a mathematical representation of the typical course of learning.

Given learning models of sufficient sophistication, there may ultimately be mechanisms of capturing and making use of variations in individual learning. Yen (1978) suggested that information processing models of learning could enable the measurement of parameters beyond those currently in use which might be useful in diagnosis of learning difficulties and prescription of alternate training approaches. Fitting of individual curves with novel (or unknown) parameters is methodologically complicated, and the return in training value

of unknown utility. Cronbach (1975) moved away from his earlier advocacy of ATI-based training, believing that the interactions between aptitude and instruction method were too complex and too task specific to be satisfactorily determined.

Different Training Methods

A substantial part of the literature on learning curves involves the comparison of acquisition performance under two or more alternative instructional or training methods. Although the effect on acquisition shape of an alternative method varies in general as a function of the type of task and its difficulty, a typical finding is that the method being used as the current or "baseline" approach tends to show the traditional negatively accelerated fit and the "new" or different method does not. (For example, see Sheppard's (1985) comparison of part vs. whole task practice on a simulator). Curves under "better" or enhanced training conditions tend to start at a higher level, rise more steeply, and frequently show higher asymptotes, often with a shape that does not correspond to any of the conventional curves. These curves behave in general as if the new method both made the task easier (increased rate) and increased the benefit derived from prior experience (higher initial value). Both these effects determine the overall shape of the curve.

Such an outcome creates an interesting paradox. If the new enhanced method were to become the baseline in another study, compared to a still "better" method, it might be expected to assume the conventional form to which the new method would be superior. Such consistent findings suggest that factors might be operating to influence learning outcomes in favor of one or the other of the methods being compared. These factors are likely to be subtle and not obvious to the experimenter. Conway & Schultz (1959) warn against the hazards of looking for a particular outcome in studies of curve characteristics,

suggesting that observed data can be influenced by an expected result, particularly in institutional settings.

GROUP VS. INDIVIDUAL CURVES

For almost 90 years, investigators have collected learning data in an attempt to understand the processes by which skill is acquired. These investigations have almost universally used group performance to represent changes in learning with practice, and have, almost since their inception, been criticized for doing so. There is validity in many of these criticisms, as well as some failure to distinguish between correct and incorrect (or valid and invalid) uses of group and individual data. Group curves and individual curves are both useful, but are appropriate for different purposes, just as are any average measure and the scores which compose it.

Characteristics of Group Curves

Average or group curves reflect the pooled performance of all group members on each trial. They represent the algebraic midpoint of performance for those individuals measured. As such they provide the best (and the appropriate) estimate of performance expected from similar individuals not included in the measured group, as does any mean value. Individual performances are clearly not useful for such a purpose. In representing average performance, group values benefit in stability from being based on multiple measures. They are thus the most reliable and informative way to describe overall group standing on the performance variable as a function of time, and together with their associated variabilities, are the most generalizable to future samples of similar individuals or to future performance of the measured group of any information available. Group curves are not, on the other hand, useful in any meaningful sense for description or representation of the learning performance of individuals. For any context in which

the focus is on the shape, pattern or rate of individual learning progress, the use of group curves is inappropriate and can be misleading, unless individual curves are shown to take the same general form. This is particularly important in evaluating performance to criterion, since a few extremely high scores (for example) can mask the failure of a large part of the group to achieve desired performance.

Characteristics of Individual Curves

An individual curve describes only the specific idiosyncratic learning pattern of a single trainee. Because it is based on only a single measure at each point, an individual's curve is likely to be unstable and erratic across time due to measurement error, even when its basic form is orderly, and such instability and variation can obscure any underlying functional shape that may be present. Instability of performance data is particularly prevalent in early trials and may continue in individual performances for extended periods (Bittner et al., 1984; Carter & Woldstad, 1985).

As with any estimates, an individual curve can be used to represent the group curve of which it is a part (although with considerable error), but a group curve cannot logically be generalized to any of its components. This asymmetry in generalization has led to some insightful analyses of the inferential risks possible when such generalizations are attempted.

Problems in Generalization

Estes (1956) points out that average curves can either over or underrepresent the level of individual acquisition, particularly prior to the stabilization of individual curves. Hayes (1953), in his discussion of criterion-based curves, terms group learning curves "irrelevant" for the study of learning,

since the form of the group learning function is determined by both the form of individual curves and the distributions of their parameters, and the same group curve can result from many quite different basic shapes for the individual functions.

Baloff and Becker (1967) show that average curves can not only obscure but seriously misrepresent the basic shape of acquisition performance. They extend Hayes' argument on individual curve shape interacting with rate and level parameters, and provide sets of theoretical outcomes in which shape of a group curve can be both reliable and completely rational and still completely different in form from the basic shape taken by every component individual curve. They demonstrate a series of individual "all-or-none" (insight) learning curves which summate to a smooth increasing negatively accelerated group curve (similar to power function shape). Summation of individual convex curves produced a concave group curve. Similarly, summation of exponentials yielded a sigmoid group shape, and summation of linears resulted in an exponential shape. While some of these arrangements are extreme cases, they nonetheless reinforce the risks involved in uncritical averaging of individual performances for purposes of inferring rate and shape of learning.

In most applications, individual and group curves rarely differ as dramatically from one another as the outcomes shown by Baloff and Becker. Both Newell and Rosenbloom (1981) and Mazur and Hastie (1978) reported some individual curves fit with a variety of functions. Their data (and that of others) suggest that, although individual curves are fit more poorly by all functions than their associated group curves, dramatic departures from group shape (completely different functional form) are rare.

Factors in Discrepancies

In general, where individual curves are reported in the literature (far too seldom), they tend to have the same overall shape as the group curve, but behave more erratically and unpredictably from trial to trial, due both to individual instability and to measurement error, and the rate, initial value and asymptote parameters of individual curves are likely to vary somewhat from the group parameters.

Although there are sound technical reasons for expecting discrepancies between group and individual curves, group curves are not necessarily poor representations of individual acquisition patterns. There are a variety of factors which can act to increase or decrease the discrepancies between the two, primarily variation in the characteristics of the task and variables which control the extent of variability in the trainee population. Virtually all the task characteristics previously identified that affect curve shape, particularly task difficulty and task complexity, also affect differences between group and individual curves. The more difficult the task, the more likely it is that trainees will show different learning rates. Likewise, a multicomponent task which requires development of strategy or extensive integration will result in greater shape differences among individuals and thus increase divergence from group patterns. Greater departures will also be present when the trainee group is heterogeneous with respect to initial ability, prior experience and ability to learn. Each of these increases variability in group performance and is likely to widen differences among individuals in acquisition rate and level.

It is thus not always the case that group curves misrepresent individual learning. Under some circumstances, some generalizations of group to individual may be sufficiently accurate for representation of learning with a single curve.

The risk is that they can not be counted on to be accurate, nor is the degree of generalization determinable in the individual case until after acquisition is complete.

Resolving the Controversy

The choice between "Group vs. Individual" curves must be decided on the basis of the purpose for which measures are intended. When it is desired to establish a "criterion" progress curve to which the performance growth of the "typical" trainee should conform, group curves are clearly appropriate and indeed required. They serve as a standard against which individual progress curves can be compared to determine the need for "intervention," for diagnosis and remediation of difficulties, or to suggest alternative ways of instruction for an individual trainee.

Such group curves are also likely to be of value in establishing a "mastery level" for a body of material which each trainee must attain, and are essential for comparing performance on the same task under alternative methods of instruction. For those applications, it is necessary to deal with an additional "parameter" of the group curve - the percentage of the group attaining criterion at each trial in the series. This quantity is not addressed in conventional analyses, and must be known in order to provide a complete picture of group progress or to serve as a basis for termination of training. Thus, for purposes of developing and operating a training program, both group and individual data are required, but sometimes with special ways of examining or summarizing the data.

If the goal of data collection and curve analysis is to gain understanding and insight about the basic processes underlying acquisition of skill, many of the criticisms of group curve generalization become more valid than if the goal is training program development. Few, if any, individuals in a group will

follow exactly the group learning function. Most will depart from it, and some will deviate significantly. Although, as we have noted, some of those departures are random, others represent real differences in how and at what rate learning is taking place. Discrepancies of individual performance from the "expected" pattern may then be resulting from real differences in learning strategy, ability, interactions with method of instruction and a host of other factors which are important for understanding the learning process and should be addressed. Thus the "unit of analysis" for theoretical or explanatory studies is the individual learning function. The summation of those functions, the group curve, may be useful in describing the general trend of performance, since its average points are relatively stable and represent reliably the broad shape of the function. In understanding learning, however, individual differences must be dealt with along with the regularities, and the group function can both misrepresent and obscure the interplay of individuals with the learning task.

BEHAVIORAL VS. ENGINEERING APPROACHES

We presented previously a variety of mathematical equations proposed for describing changes in performance of a task as a function of practice or experience. These equations, while they showed substantive similarities in form, were brought together from two rather diverse sources. The principal studies of skill acquisition from the standpoint of human learning have come from disciplines with a "behavioral" orientation (primarily experimental and cognitive psychology). Curves and descriptions of acquisition patterns based on the same mathematical approaches but viewed from a somewhat different orientation have been used in engineering, primarily industrial engineering, for almost 50 years. Developments in the engineering community have taken largely independent tracks from those in the behavioral literature (with a few crossovers). Engineering analyses have traditionally approached improvements in task performance as an

increase in productivity with output or "cycle time" as the basic metric, and have shown little interest in the processes underlying these improvements or in the changes taking place in the individual, in contrast to behavioral efforts which have taken the understanding of such processes and changes as their principal goal.

As a consequences of these different orientations, there are substantial differences between the two disciplines in the kinds of tasks on which data are collected, the time periods involved and the application of descriptive functions. In general, the behavioral and engineering literatures can be distinguished on the basis of five factors.

Nature of Tasks Used

Curves from the engineering literature are almost without exception based on real production tasks. The measure is typically output per unit time and is obtained in a job setting as part of the in-place production system. As such, measures are usually representative of system output, and involve variations due to aspects of production (equipment, logistics) unrelated to operator capability per se. Tasks are rarely constructed (i. e., experimental) and invariably multicomponent, involving (usually) integration of motor, procedural and planning skills. Task output is likely to have inherent upper limits brought about by equipment or other task pacing constraints. These produce artificial asymptotes that occur earlier in practice and at a lower level than in tasks without such limits. This in turn causes differences in the shape and parameters of the curve. DeJong (1957) refers to this phenomenon as "compressibility" and suggests corrections to the functional equations to adjust for its effects.

Emphasis on Group Output

Both disciplines have relied on group performance description, much more so in the engineering fields. Many of the industrial applications involve using the "typical" curve of performance growth (often of predetermined shape) to estimate the time course of production prior to "gearing up" for a production process. Individual deviations from the typical curve, so long as group productivity is as expected, are not of great significance unless the lower performance continues over extended time frames.

Very Long Time Periods

The periods of time over which performance improvement is expected is typically months or years, rather than the days or weeks for experimental studies. Tasks on the whole require greater integration and continue to show growth for extended practice or experience times (see Taylor & Smith, 1956). This is characteristic of the learning patterns encountered in real-world tasks, in which improvements imperceptible over momentary periods continue to accumulate long past the ordinary periods of "learning" or training.

Motivation and Interest of Trainees

Because of the economic incentives inherent in the industrial situation, it is presumed, in most cases realistically, that trainees and/or operators are highly motivated to perform, since their continued employment can depend on a satisfactory learning rate and ultimate productivity. For much of the behavioral experimentation, this strong interest in maintaining output may not be present. Subjects are less likely to exert continued effort for many repetitions of a task over the extended sessions required for performance to stabilize and level off. Effects of this

flagging motivation were demonstrated in a recent study involving 15 repeated administrations of a series of tests over several weeks (Kennedy, Dunlap, Wilkes & Lane, 1985). After showing conventional growth curves over the first 7 or 8 practice sessions, group data past the 10th administration developed substantial irregularities (apparently random increases and decreases in performance), and some individual curves showed aberrations as early as the 6th or 7th session of task repetition. These effects were attributed to boredom and a decreased concern with maintaining performance levels. Motivation in an on-job setting is likely to be both higher initially and better sustained over time.

Handling of Poor Performers

Related to the issue of trainee motivation is the extent to which individuals may be dropped or separated from the study (in experiments) or from employment (in industry) prior to conclusion of performance tracking. It is clear in analyses from the industrial engineering literature that persons who fail to show some acceptable level of performance growth are likely to be dropped from training or to have their employment involuntarily terminated. This is presumably a less common event in the collection of data from an experimental cohort. The probable outcome of such differential retention is that performance is higher (and likely more predictable) in the industrial setting since 1) motivation is enhanced, and 2) poorer performers are dropped from the industrial population at some point in the time course.

Industry makes extensive use of production curves for a variety of purposes. Nanda (1977) describes seven typical applications of the "learning curve" or manufacturing progress function in industrial operations: Cost estimating, scheduling, efficiency comparison, procurement and subcontracting, personnel planning, facilities and long-range planning. It is of interest

that training planning and evaluation is not specifically spelled out, but such concerns are addressed in depth elsewhere (Glover, 1966; Hancock, 1971; Levy, 1965).

It is important to recall that the "learning curve" concept as used by Nanda and others does not necessarily include any substantive component due to differences among operators or trainees or any contribution to total performance growth as a result of improvement of the operators. Depending on the nature of the process involved, all or most of the output may be a function of improved logistics, more effective management, and other support aspects of the system itself "learning" to produce more efficiently with greater experience. Conway and Schultz (1959), for example, believe that operator learning as a contribution to increased output is "overemphasized" and may be of "negligible importance" in many manufacturing situations. Given such different definitions of "learning," it is remarkable that curves from behavioral and engineering studies are so highly similar in form.

These inherent differences between behavioral and engineering studies have implications for the generalization of findings to military training situations. The emphasis on group training, the sustained periods of task performance, the higher levels of motivation, and the conduct of training on real tasks seen in industrial settings are all more characteristic of the military training environment than are the conditions under which the typical behavioral experiment is carried out. In some military training (particularly that after basic), the dropping of trainees for non-performance or their treatment with special or additional training is much like that in the industrial sector. As we will discuss later, the conclusions and generalizations from the various literatures do not differ in any dramatic way. The principal differentiation comes with respect to a) the extensive history of practical applications of curve shape and level in industry as a standard against which

performance of the production system or of individuals can be compared, and b) the willingness, for economic reasons, to maintain records of both and group individual output. These applications are close in spirit to our previously stated goal of using performance curves for determination of appropriate time courses for segments of military training.

SOME RELEVANT THEORY AND FINDINGS ON ACQUISITION

A major objective of the preceding sections has been to draw from the literature on acquisition curves some indications of where and how military training can be improved, particularly in the domains of course and segment duration, pacing, scheduling, and other areas that involve decisions about training intervention or termination. Even though the concern is primarily how to improve training procedures, such decisions inevitably impact the effectiveness and the cost of particular training arrangements. We have looked also for regularities and predictabilities in form and rate of acquisition with a view to extracting some recommendations oriented toward those domains, with a passing interest in retention as it relates to acquisition variables. It is possible, from the information in previous sections, to derive a few such recommendations on the basis of empirical outcomes alone. There are some reasonable regularities in the literature with respect to expectations under certain conditions of skill acquisition, but there is still considerable uncertainty about why these regularities occur. The value of recommendations and likely their generalizability would be enhanced by some indications of how they relate to the more theoretical underpinnings of learning and acquisition.

It is not our intention to review all or even any major part of the acquisition (or retention) literature. Both literatures are extensive and complex, and it would be redundant to present and discuss their subtleties in-depth (since so many others have done so), and also unnecessary, since we are only concerned here with those aspects of acquisition and retention that might dictate differing approaches to training. Where the data are in concurrence about the expected outcome and how to bring that about, theoretical explanations underlying the outcome will be of less interest. We will thus make extensive use of previous

reviews and syntheses, in addition to a number of original sources, and will only comment or expand on major issues or highlights that are germane to our main thrusts.

There have been a number of analytic summaries of the literature related to acquisition and learning. Some are quite current, others are primarily historical. Some deal with almost exclusively theoretical issues, others use a framework closely allied to our interest in military training and education. We have referred previously to articles in the Annual Review of Psychology by Glaser and Resnick (1972), McKeachie (1974), Resnick (1981), Gagne and Dick (1983) and Wexley (1984). Other reviews or extensive summaries of more delimited areas contributed both background information and detailed findings. These are identified below by their major area of content; we will also note them as appropriate in later discussions.

Acquisition and learning -- Bilodeau and Bilodeau (1961), primarily motor skills; Irion (1966), skilled performance with emphasis on motor learning; Powers (1976), individual differences in instructional effects; Kleinman (1977), motor learning; Horton and Mills (1984), learning and memory processes, including retention; Salmoni, Schmidt and Walter (1984) on motor and perceptual learning with emphasis on knowledge of results; Adams (1985), skilled performance and motor learning; Kulhavy (1977) on feedback in academic instruction; Reigeluth, et al. (1982), instructional theories and design; and Dansereau, Atkinson, Long and McDonald (1974) on the formation of learning strategies. For cognitively-based theories of learning (in addition to Gagne and Dick (1983)), Anderson (1982) and the majority of chapters in Anderson (1981).

Transfer of training -- Naylor (1962), part-task training; Valverde (1973) and Spears (1983), predominantly in the flight training domain; the unusually thorough review in the military context by Cormier (1984), which covers most previous summaries; Adams (1985), motor skills.

Retention (as it relates to acquisition issues) -- Naylor and Briggs (1961), skilled performance, primarily motor; Underwood (1964), verbal learning; Gardlin and Sitterley (1972), highly skilled performance including procedural components; Leonard, Wheaton and Cohen (1976), general retention and transfer with military task emphasis; Prophet (1976), flight skills, predominantly military, and Smith (1976) in a similar vein; Taylor and Thalman (1977), general military tasks; Schendel, Shields and Katz (1978), extensive review and analysis of motor skills retention in military context; Slamecka and McElree (1983), verbal material; Hagman and Rose (1983), summarizing a series of studies on retention of (mostly) procedural tasks in the military, and Farr (1986), for a review and analysis of processes underlying long-term retention.

The theme that runs through this paper is an emphasis on performance, particularly skilled performance on realistic tasks. It is our belief that skill acquisition (including the learning and application of declarative or factual material) proceeds in accordance with a set of stages or phases of learning. While the trainee may appear to be repeating or practicing the same task throughout the learning period, the nature of the underlying changes occurring in skill development differs among these various phases. Changes in the unobservable infrastructure of learning lead directly to the differential effects of practice on performance across time, and indirectly to the changes in rate and shape observed in the acquisition curve. One key to understanding the use of learning curve data is thus the general nature and effect of various stages in learning as an individual acquires a significant skill.

"Skill" is an evasive concept. Like so many terms in the behavioral literature, it has been used by different writers with a diversity of different meanings. These variations have been relatively wide ranging and have evolved steadily over

time. Adams (1985) gives a historical perspective on the changing meaning of "skill" over the last hundred years, and notes the confusions engendered in research on skill without explicit attention to the definition of the concept. Such a consideration of meaning is, in our judgment, necessary for the treatment of two basic theoretical issues. One deals with the characteristics of behaviors subsumed under the term "skilled performance." The second is concerned with differentiating between skill and ability, in particular the extent to which abilities are viewed as permanent attributes or as attributes modifiable by practice and experience.

THE NATURE OF SKILLED PERFORMANCE

Different investigators have studied "skilled" performance with a range of tasks from those with simple "purely" motor requirements to those which deal with already developed ability and require little or no learning. Usage of the term has moved over the last decade to reflect an interest in skill learning and to be progressively more constrained to tasks which impose relatively high demands on the learner. From our review, it is clear in present usage that a) "skilled performance" refers to the acquisition of a learned capability to perform tasks of much greater complexity than the traditional laboratory task and b) such highly demanding tasks have become overwhelmingly the main topic of interest in studies subsumed under the general rubrics of "skill acquisition" and "human performance" in current research.

Despite some minor differences, most recent definitions of skilled performance show substantive agreement across diverse domain areas on the essentials of skilled behavior. In general:

a) Acceptable performance requires "at least 100 hours" of training or practice (Anderson, 1982; Schneider, 1984). Simple tasks and short time frames are insufficient for realistic study

of skilled performance. Skill develops over thousands of repetitions (versus the hundreds for typical laboratory tasks), and may continue to improve across extremely long periods of practice (Annett, 1971; Ehrlich, 1943; Schneider, 1985), even, for some tasks, tens of thousands of trials (Newell & Rosenbloom, 1981). The gradual reduction in improvement per trial with practice tends toward a negatively accelerated performance curve.

b) Skill is a within-subject, not a between-subject, phenomenon (Bilodeau & Bilodeau, 1961), and changes in the individual are a key unit of analysis. Initial performance is highly unstable (Ehrlich, 1943; Schneider, 1985), and is characterized by wide individual differences in rate of learning. Gettinger and White (1979) report ratios of time to learn between fastest and slowest students in the range of 6:1 to as high as 13:1 for some materials. At least some of the subjects never become truly proficient despite extended practice (Schneider, 1985), and there are typically attritions from formal training or where preestablished standards exist (Glover, 1966).

c) Skilled tasks are multi-component and heterogeneous in nature, requiring mixtures of cognitive, motor and perceptual abilities (Adams, 1985; Schneider, 1985). Acquisition through learning of performance strategies is typically necessary (Dansereau, et al., 1974), requiring lengthy periods to resolve and creating false "asymptotes," "plateaus," and day to day "dithering" of the performance curve before it stabilizes (Glover, 1966; Kao, 1937; Schneider, 1985; Taylor & Smith, 1956).

d) Highly skilled performance is characterized by distinct qualitative differences between expert and novice. Krendel and Bloom (1963) define three characteristics of the "natural" or extremely proficient pilot: 1) Economy of effort (far less energy and attention is required than for the less expert

pilot); 2) Consistency of performance (goal-related output is constant for many different conditions of input); 2) Adaptability (performance mechanisms are automatically adjusted to compensate for wide variations in task conditions or to maintain performance in information-poor or reduced feedback environments). The latter characteristics correspond to Spears' (1983) description of skilled performance as "robust." Acquired skills are adaptable to a variety of situations in which cues are modified or distorted and response outputs must be correspondingly changed to accomplish objectives. The mechanism underlying adaptability and economy is what has become known as "automaticity" or "automatism" of behavior (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977), which has gone beyond its earlier origins in (primarily) motor learning to emphasis on cognitive and integrative skills (Neves & Anderson, 1981; Schneider & Shiffrin, 1985; Shiffrin & Dumais, 1981).

DISTINCTIONS BETWEEN SKILLS AND ABILITIES

Skilled behavior is an observable. Task-related activities of the learner can be determined and recorded without the imposition of theory. The existence of data (although not necessarily its meaning) is directly verifiable by objective means and does not require inference. Skill, on the other hand, is a construct whose existence and characteristics, like that of any other construct, must be inferred from the properties of data. Skill, as a term, tends to have two general meanings in use. One usage describes a constellation of behaviors that make up "a skill." In this context one refers to, for example, "flying skill," and the word "skill" has a behaviorally anchored meaning independent of any particular individual. In its second usage, it describes something an individual possesses, a capability to perform a task with a high order of proficiency. An individual can have more or less of this capability and can improve it with practice. In this latter usage as a capability, skill is extremely difficult to distinguish from the traditional

meanings of "ability," except in the degree of modifiability it is allowed by whatever definition is applied.

Fixed Abilities, Changing Tasks Models

For a number of years the distinction between skills and abilities received considerable attention in the research literature, mostly in regard to issues of modifiability. A classic and recurring finding in studies of practice is that correlations between successive performances of the same task tend to decrease systematically as the number of intervening trials increase. In this pattern, referred to as the simplex (or in the more general case, as superdiagonal form) (Humphreys, 1960; Jones, 1959, 1960), the predictability of one trial outcome from another gets less as the trials are further apart; typically, performance in late practice is predictable from neither early practice or any other data (such as ability measures) available at the outset of training. This finding has been repeatedly replicated (e. g., Reynolds, 1952b). It strongly suggests that something is changing about either the individual or the task (or both, an explanation that was not considered acceptable for a long time). This correlational pattern was studied extensively by Woodrow (1938, 1940). Woodrow, presuming abilities as fixed attributes of the individual, necessarily interpreted outcomes as indicating that the pattern of abilities required for good performance must be changing with practice. In this view, ability is the capability (in the sense of potential) to do something; learning of a skill is a matter of rearranging or combining those abilities (Woodrow, 1946).

The fixed-abilities viewpoint has received considerable empirical examination. Fleishman's factor analytic work, conducted over a number of years (Fleishman, 1966, 1972), looked at shifts in the factor structure of performance as a function of practice. He found that different factors were important in

predicting performance at different stages of practice. For Fleishman, abilities were relatively enduring, general traits of the individual. While they were a product of learning and experience as well as of any innate aptitude, abilities applied to a variety of task types, serving as the base with which the learner started in new learning situations. Skill, in this view, is task-specific, representing the level of proficiency attained on a given task. Abilities, while they could be modified by learning, were basic attributes, much fewer in number than task types. They, like skills, are abstractions or inferred concepts.

Fleishman's use of factor analysis on practice data, and the resultant conclusions about changes in factor structures, have been criticized by Humphreys (1960), among others, and implicitly by Jones (1960, 1970), who offers alternative structural models based on molar correlational analysis. Jones' interpretation, however, paralleled Fleishman's in its contention that different abilities are important at different stages (although the factors are very differently defined). He also offered in his two-process theory (Jones, 1969, 1970) a mechanism which provides for both the observed regularities in learning data (which suggest a lack of individual differences) and the idiosyncratic learning strategies or style invariably seen in complex tasks. The two-process theory suggests that complex tasks are composed of simpler ones, and that learning proceeds differently for simple and complex tasks. In the "terminal process," simple tasks become simpler with practice (different and/or fewer factors involved), performance becomes essentially perfect, and individual differences gradually disappear, leaving only error variance (movement toward "automaticity"). Complex tasks are made up of combinations of simpler tasks, but tied together in ways and with strategies that may reflect a particular individual's learning style, and his learning rate (the "rate" process).

Jones' two-process theory, Fleishman's (1972) analyses, and the bulk of Adams' work (summarized in Adams, 1985) produce, from different theoretical positions, similar conclusions about the behavioral content of skilled performance after extensive practice. Task performance becomes progressively more task specific with increased repetitions; its interrelationships with variables other than itself systematically decrease, and it correlates well only with recent trials on the same task until performance stabilizes and approaches asymptote (Jones, 1985). In his most recent analysis, Jones (1985) uses this property of task intercorrelations to track the point at which additional practice no longer enhances retention. Once intertrial correlations are stable and high (the curve has leveled off), the trainee is in a period of overlearning. Jones' analyses to date suggest (for the tasks employed) that no major gains in retention occur after four or so trials of overlearning, even though training performance may be continuing to improve slightly. Such a finding offers the potential to determine empirically the appropriate number of trials past stabilization for a given training task and to use the onset of relatively high, stable correlations as a tracking mechanism for when to terminate training.

We noted earlier that skilled behavior can be defined and observed external to the individual or task situation, whereas all the above viewpoints require the existence of two distinct entities to determine that behavior -- abilities and skills. Both are constructs inferrable only from external behaviors; both are to some extent modifiable by experience or learning (although abilities are seen as only slightly so). It is easy to see in retrospect the extraordinary difficulty of distinguishing either rationally or empirically between two internal constructs so similar in effect, regardless of how divergent they may be in theoretical meaning.

A major weakness of the "changing-tasks" model for explaining intertrial relationships is thus the logical near-

impossibility of a) identifying fixed or only slightly immutable abilities that occur in many different task circumstances, and b) further demonstrating that it is their variation in importance across trials that causes the relationships, and then c) showing how such an assumption about abilities can be differentiated from a model that simply assumes that practice changes the strength of the abilities required to perform the task. Adams (1957), in an early criticism of changing-task approaches, noted that his data was explained as well by one model as by the other. Humphreys (1960) makes the same point; a measure samples current capability no matter what it is called. Kleinman (1977) also gives a comprehensive summary of the issues involved in ability-skill differentiations.

Changing Abilities Models

While it is clear that, with practice, tasks change in their relationships to earlier performance and to other variables, becoming more specific across trials, there are several important weaknesses in the explanation of these changes by the presumption of fixed or only slightly modifiable abilities. First, it is difficult for the reasons previously stated to make logical distinctions between the two constructs of skills and abilities. Second, the lack of modifiability of abilities is inconsistent with an entire literature on transfer which holds that learning of a given task is enhanced by practice on a similar task. If it is considered that each new task is approached with the same set of abilities as any previous task, that only learned skills transfer, and that such skills are task specific, it is then necessary to posit as many starting positions for each individual as there are combinations of previous tasks for him to have learned. This is far from a parsimonious position with respect to understanding a trainee's initial status on a new task.

The same outcomes which are explained by the changing-task, fixed-ability presumption are also consistent with a model that

allows abilities to change with practice. In such a view, the decreasing relationships across trials do not necessarily imply changes in ability requirements of tasks, but can be due to modification of abilities such that the abilities used on tasks in late practice are no longer identical to those in early practice. In a landmark article, Alvares and Hulin (1972), showed that the instability of early practice (and by extension the decreasing trial interrelationships) could be attributed to subjects changing abilities at different rates across practice, and that the presumption of different combinations of fixed abilities at different stages of practice was unnecessary. In a later article (Hulin & Alvares, 1973), they addressed the tendency of late performance to be unrelated to virtually all measures available at the start of practice, and concluded again that this phenomenon could equally well be due to changes within the subjects rather than to changes in the importance of particular abilities across stages. They further reasoned that the question of changing-tasks vs. changing-abilities was difficult to resolve either empirically or logically, and suggested that the distinction be abandoned. This suggestion seems to have been followed in more recent work (see Adams, 1985).

Implications

For the purposes of the present analysis, skill-ability distinctions are not critical. The increasing task specificity and the early instability of practice involved in these distinctions are, however, highly germane to our concern with decisions about the appropriate place to terminate a segment or module of training. We noted earlier Jones' (1985) finding that termination prior to the period of relatively stable performance adversely affects retention. Stopping practice while a trainee is still in the "steep" part of the learning curve (before the curve levels off) causes retention over time to be reduced disproportionately to the trials remaining before

stabilization. It is thus important to be able to tell where on the curve an individual (or a group, if curves are representative) is located in order a) to ascertain whether training in a given segment is essentially complete, and b) to avoid overlearning that will not yield sufficient gains in retention to justify its cost. Further, as task specificity increases, less and less of the gains of learning will transfer to a new skill and, depending on task similarity, negative transfer could eventually be induced by too long a practice period.

STAGES AND PHASES OF LEARNING

There is little disagreement in the literature that learning of a complex task (and probably most simple tasks) proceeds in accordance with approximate segments of practice. Within a segment (henceforth a "stage"), both overt activities of a learner and the internal changes occurring are different in a qualitative way from activities during other stages. The existence of stages in some form is supported by virtually all the research and theoretical development of the last two decades. The presence and the nature of stages of learning have considerable implications for examining the time course of training and in making decisions about remediation, intervention and termination.

The number of stages and the labels that are appended to each stage are obviously a matter of individual theorists' choice and are to some degree arbitrary. Throughout our review, however, we encountered a surprising agreement across an extraordinary breadth of domains on the number of stages, and a remarkable correspondence between the content and description of activities occurring in each stage. The magic number of stages is three. There are further some very sensible relationships between the nature of stage models and the general patterns of acquisition curves observed.

The skill development stages we will use are those of Paul Fitts, originally postulated (with slightly different names) in Fitts (1962) and refined and expanded in Fitts (1964) and Fitts and Posner (1967). These have been used widely throughout the literature, and have served as the basis for other similar stage developments. We will map into these stages some categorizations employed by other investigators, and describe some relationships between the stages and our analyses of acquisition curves.

Fitts' Stages

The Fitts (1964) stages for skill acquisition (along with some of our own observations) are:

a) Cognitive -- In this stage, the learner exerts effort to understand the task to be performed. There is an initial encoding of the skill into some primitive form sufficient to generate responses which approximate in a rough way the desired behaviors. The "rules" are learned, and strategies for approach to the task are developed and tested. Practice on individual task components is conducted one at a time and without integration. Fitts indicates the typical presence of overt verbal mediation or rehearsal of specific information required for task performance. This stage corresponds to early practice, and performance may be characterized by considerable instability as strategies are tested and discarded and emphasis is placed on separate components of the task. Depending on task difficulty and the degree of prior experience of the learner, this can be a period of either very slow growth or extremely rapid growth (if difficulty is low or moderate or prior experience is high). The learner transitions from this stage with a basic understanding of task requirements and rules and a set of strategies for successful performance, not yet fully elaborated, encoded or implemented.

b) Associative -- In this stage, the skill is "smoothed out." Deficiencies and errors in initial understanding of the task are systematically eliminated and strategies are refined. The appropriate stimulus-response links are established, and preliminary motor programs are developed. Rudimentary integration of task component skills is initiated and whole-task practice has begun. Dependence on verbal mediation is eliminated. Performance is still primarily under voluntary control and attention investment is high. Progress occurs rapidly and the learning curve is typically very steep during most of the stage. By the end of the stage, increments to performance between trials have begun to decrease and the performance curve has "turned the corner" and started to level off as initial asymptote is approached.

c) Autonomous -- This stage, which occupies (at least potentially) the longest practice periods in development of highly skilled performance, is one of gradual improvements over long sequences of repetition of the task. Improvement, as we have noted earlier, can continue for long periods, spanning many thousands of trials. Performance control programs developed in the associative stage are systematically refined, and component integration is complete. Practice serves to shift control from overtly voluntary processes to low effort "automatic" control of performance.

The Fitts' stages or equivalents, although developed from a predominantly "motor" orientation toward skill development, have been observed to hold over a wide range of task types, including those with significant "cognitive" components involving little or no motor requirement. Despite their appealing logic, the segmentation of skill acquisition into stages has been like the "ubiquitous" power law - both recur in diverse applications and show high and regular correspondence with the data, but the analytic rationale for why they work has been slow to develop.

Anderson's Stages

The most powerful theoretical analysis has been that of Anderson (1982), working from the concept of cognitive activity. Interestingly, Anderson's development addresses the micro-activities underlying the stages of acquisition while producing an explanation of the power law as a by-product. Anderson, taking Fitts' stages as descriptive but not explanatory of the course of learning, posits a set of basic learning processes to account for the commonly encountered stages. Even a summary of Anderson's mathematically elegant reasoning is well beyond the scope of our survey, but the following describes in brief the stages used by Anderson and their relation to those of Fitts.

Anderson's theory is based on acquisition of skill by the learner in the form of a production system. As is characteristic of cognitive theorists, he distinguishes between procedural knowledge (information) which is represented in the learner by production rules, and declarative knowledge which is cast in a propositional network of facts which are operated on by the productions. The similarity between Anderson's terminology and that of Gagne (1977) and other educational theorists summarized by Reigeluth, et al. (1982) is noteworthy. Anderson's developments provide an effective rationale for many of these theories as well as for the Fitts' stages. Following Anderson's (1982) presentation, his stages are:

a) Declarative -- The learner receives facts, information, background knowledge and general instruction about a skill. These facts are used by general procedures already possessed by the individual to generate an approximation of appropriate behavior. Verbal rehearsal is high because facts, being new to the learner, must be kept in working memory to be available for use. Activities are analogous to those in Fitts' cognitive stage.

b) Knowledge compilation -- Practice causes basic knowledge and facts about the skill to convert gradually from declarative form into appropriate new procedures which can be applied directly to the processing of inputs without constant voluntary attention. By the end of this stage domain knowledge has been compiled into a set of production rules for linking input to output. Anderson considers this stage (Fitts' associative) as an intermediate or transition state between the declarative and procedural stages in his framework.

c) Procedural -- After declarative knowledge is compiled into a production system, practice refines and strengthens appropriate procedures; they show both generalization to similar tasks and discrimination (specialization for specific task situations). The strength of procedures "accumulates." The accumulation of response tendencies in underlying processes generates steady improvement in performance over time (in accordance with the power law). Neves and Anderson (1981) expand on the steps by which procedures are refined or "fine tuned" with practice. They suggest processes of Composition, in which parts of the total procedure are grouped so that fewer procedural activities must be initiated; Speed-up as a result of step reduction through composition and through elimination of unnecessary steps; and Automaticity, in which processing of some steps is conducted in parallel vs. sequential.

Anderson's system focuses primarily on the nature of activities in the Declarative and Procedural stages, treating Knowledge Compilation as a process by which learning moves from one of these principal stages to another. His mathematical development describes the steps by which production rules combine into a complete, integrated performance management structure. Among the key aspects of the development is the nature of competition among alternative production rules. Several different rules may be effective in generating appropriate responses. In Anderson's model one rule does not

necessarily replace another; alternative rules may coexist, with one or another becoming gradually more strengthened as it is found through experience to be more accurate or more efficient. As we noted previously, this has important implications for the expected form of the acquisition curve. As Mazur and Hastie (1978) note, accumulation models yield the hyperbola (a power law), while replacement models predict the exponential function.

The production-system approach to study of acquisition described by Anderson has been used successfully by other cognitively-oriented investigators. Kieras (1985) and Kieras and Bovair (1985), for example, use a production system analysis of the acquisition and transfer of procedural information acquired from written instructions in textual format.

Rasmussen's Paradigm

Rasmussen (1979) describes three categories of skilled behavior. Although Rasmussen is in part concerned with describing the nature of task contents for the categories and their implications for design and operation of equipment and for training, the three categories show a remarkable compatibility to the stage models just discussed. Categories are:

a) Skill-based tasks (or task components) are composed of simple stimulus-response behaviors, well-learned by repeated practice, highly automated and requiring minimum conscious control. The rules which link stimulus and response are straightforward, requiring no decisions or interpretations, and are of the form of one specific input cue requiring a single, unambiguous response to produce the desired output (if the red light comes on, press blue button). Rules are provided to the operator (or trainee in our terminology) and practiced until they can be efficiently and rapidly executed.

b) Rule-based behavior involves the ability to recognize specific, well-defined situations and choose the appropriate

procedure for response based on unambiguous decision rules provided to the learner. The operator discriminates situations which call for one rule rather than the other, and to master rule application, but does not generate or change rules. Rule-based tasks are likely to represent combinations of skill-based behaviors combined in rule-based ways (if the system is in mode y, and gauge A exceeds 100 and gauge B is less than 10, press button x).

c) Knowledge-based skills are used in situations for which familiar unambiguous cues are absent and clear and definite rules do not always exist. Tasks involve both discrimination and generalization of rule-based learning. The trainee must relate the current situation to one previously encountered and synthesize or abstract rules on the basis of similarity of the situation to prior experience. Successful performance requires a thorough understanding of the task "system" and its normal and abnormal states, and the development of a consistent model for relating inputs to outputs. Most behavior of the type we have described as "skilled performance" would be subsumed under this category, as would most real world tasks of any complexity. Most training oriented to the performance of complex tasks has as its objective the formation of knowledge based behavior. As Singer and Gerson (1978) point out, the ultimate goal of any training system is to provide the trainee with the capability to generate, test and adapt his own strategies.

Similarities of Rasmussen's paradigm to the preceding stage models are obvious. All show a hierarchy of simple processes and rules individually mastered through practice and eventually integrated into a smoothly executed set of task performance procedures, which are further refined, generalized and discriminated over extended practice and continuing exposure to more instances to which the rules of performance must be applied. Rasmussen's categorization of tasks is useful in

conceptualizing the links between task content and the type of training required for proficiency on complex tasks. It also corresponds in a straightforward way to the movement of a typical military specialist through segments and modules of basic training (progressive introduction and practice of simple tasks), to on-the-job training (practice to consolidate basic skills), and back into formal advanced training (more complex job-related skills), followed by further on-job practice, with the cycle continuing throughout a career.

An additional indication of the prevalence of three-stage breakouts of the course of acquisition is the three-segment performance curve reported by Taylor and Smith (1956). As we noted earlier, they suggested that their long-term production data seemed to require three distinct curve segments, one each for early, middle and late periods of practice on the job, with the nature of activity in the segments following closely to that in the Fitts's stages.

DEVELOPMENT AND IMPLICATIONS OF "AUTOMATIC" BEHAVIOR

The evolution of task performance over practice to a stage of highly integrated "semi-voluntary" control of task activities is consistent with virtually all the theory and all the data on skill acquisition, and is particularly evident in tasks on which practice continues for extended repetitions. Although the recognition and description of the phenomenon go far back into the literature on analysis of learning (Gagne and Dick (1983) trace it to the late 19th century), the best-known current codification and expansion of the "automaticity" concept is that of Shiffrin and Schneider (1977) and Schneider and Shiffrin (1977). There has been in the last decade or so an emphasis on the learning of tasks which are more representative of real-world performance than the laboratory tasks of what Adams (1985) describes as the "Middle Period" (1940-1970). These tasks are typically multi-component in nature and require

extensive periods of practice to develop and refine strategies for component integration and for task "management" and control. Such tasks involve what Schneider (1985) terms as "high performance skills," and almost without exception show acquisition patterns consistent with the formation of automatic behavior.

Automaticity in its conventional definitions results in improved performance principally through the mechanism of increased efficiency in the use of attentional or processing resources (Schneider, 1984) or, equivalently, the systematic reduction in requirements by a task for some limited processing capacity. As practice continues, the processes for task management become so thoroughly ingrained that, once initiated, they assume the status of involuntary mechanisms and proceed virtually without the expenditure of attentional resources other than to monitor the effects of outputs on task performance.

The anchoring of automaticity development in concepts of more efficient use of capacity (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977) is one theoretical explanation of empirical findings, but not the only one which leads to the same outcomes. In the next section, we discuss some implications of schema theory (Schmidt, 1975) and its derivatives from cognitive theories. Schema theory describes the formation of automatic behavior in terms of the development with practice of "programs" for control of performance that are similar to the capacity arguments of Schneider and Shiffrin, but derived from a different theoretical base. Rumelhart and Norman (1978) propose three separate processes by which increased efficiency can occur: Accretion, which is the addition of new rules to established production systems; tuning, similar to the fine-tuning of procedures previously described, and restructuring, which involves the reorganization of procedures into more "compact" units.

Cheng (1985) also questions the Schneider and Shiffrin explanations for automaticity. She maintains that the findings ordinarily attributable to automaticity can be adequately described in terms of task restructuring. With practice, task components are coordinated, integrated and reorganized into new cognitive or motor units. Steps are eliminated or combined (as in Neves and Anderson (1981)) to form new units which are more efficient because they require many fewer procedural activities to initiate and manage the performance process. Greater discrimination is acquired through practice among the features of stimulus cues, and "filters" (our term) are developed which improve the efficiency of feature extraction in stimulus processing. (Annett (1971) offers a substantially similar development.) Cheng likens the capacity explanation of automaticity to performance improvement resulting from the "hardware" becoming faster, and the restructuring explanation to a series of "software" modifications in which, through experience, the "programs" are progressively improved to get around hardware limitations. Schneider and Shiffrin (1985), in response to Cheng's criticism, agree that both restructuring and features learning occur, and are logically part of the automaticity development process, but are insufficient to account for the totality of the phenomenon.

All the explanations posited for the existence of apparently automatic behavior are theoretically reasonable and all account, at least in part, for the observed outcomes. For our purposes, we are more concerned with the nature of those outcomes and with how to deal with their effects on the training environment.

The implications of automaticity for complex skill training are considerable. Transitions of task management from voluntary control to the "non-conscious" or involuntary mechanisms posited for automatic behavior are characterized by a segment of the learning curve in which performance shows slight but steady increases over relatively long periods of practice. It is

uncertain as to the specific point in acquisition that automaticity can be said to have occurred, but it seems clear that the development of truly "automatic" behavior cannot begin until performance has "levelled off" at initial asymptote and continued past that point for some period of practice. Retention data from Jones (1985) suggests that performance prior to the "bend" in the curve and for at least a few trials afterward is based on a qualitatively different level of skill consolidation for the learner than performance after stabilization begins. (The learner is still in the associative or knowledge compilation stage). In this reasoning, skills for which practice is terminated prior to the initiation of automaticity development would not be retained nearly as well as those for which practice is carried at least a few trials into overlearning.

A second major implication of automaticity for determining appropriate length for a training course or segment is the development of task-specific behaviors which develop under extended practice and are an integral part of the refinements which take place as task control requires progressively less voluntary attention. While retention of specific skills is likely to be enhanced almost without limit by continued practice, the degree to which any desired transfer to later segments can occur may be systematically reduced by the overlearning associated with extensive repetitions (see Farr, 1986).

Cormier (1984) reviewed and synthesized over 200 original sources as well as a number of previous reviews, emphasizing in his conclusions the generalizations of findings to the military training situation. Cormier's conclusions reinforce the risk of reduced or negative transfer associated with extensive overlearning if cues are not highly similar across segments. Studies in the review also suggest that, in decisions about the length of a training segment, the effect of negative transfer

from extended practice must be balanced against a countervailing tendency toward reduced transfer for tasks terminated before integration of task components is fully formed. In particular, the degree of task organization inherent in the task or imposed by the trainee appears to be a key variable. Both Naylor and Briggs (1971) and Gardlin and Sitterley (1972) noted that tasks with a high degree of internal organization or "coherence" are retained better than those for which organization is low. Although they are referring primarily to task organization as an inherent property of the task, it is likely that trainees impose some type of organization of their own in the absence of sufficient structure (see Morrison, 1982), and that this personal organization will become more efficient with practice.

Since a key objective of much of military training is to provide transfer of some competence to other segments and situations, tracking of the level and shape of individual trainee performance curves may be necessary to identify the "best" point for concluding training in a segment or module. Best available evidence indicates that a) practice or other exposure to material in the typical military training situation should continue for some time after "mastery" to increase retention (Schendel, Shields & Katz, 1978; Jones, 1985), but that b) such continuation should be limited in duration to the minimum trials required to attain the component integration associated with automaticity (Cormier, 1984).

SCHEMA THEORY EXPLANATIONS OF ACQUISITION

The explanation of complex skill acquisition in terms of mental representations of learning called "schemata" has derived from and is firmly rooted in what might be broadly termed the "cognitive" tradition. In the developments of Schmidt (1975, 1982), Norman, Gentner and Stevens (1976), and most other theorists, there are in memory a series of structures which contain codified information and procedures which govern the

perception, filtering and analysis of input, the processing of data and the selection of appropriate responses. These schemata form the basis for approaching all task performance situations and function for storing knowledge, understanding its meaning and that of new inputs and in determining and initiating required action.

Data structures in a schema contain both knowledge, stored in a propositional network, and concepts or rules which describe the relationships among propositions. Learning consists of processing new information into the schema, adding knowledge to the propositional network, and modifying the interrelationships among key variables to reflect inferences from the new information. A learner, observing the task environment, determines correlation and regularities among its features, forms categories into which observed "instances" are placed, and creates linkages among categories that assist both in processing additional instances and in determining appropriate actions in response to those instances. Over practice, schemata may be tuned, discarded and replaced by modified schemata, or reorganized for efficiency (restructured) into new, related schemata in a way similar to that described in our discussions of automaticity development above.

From the standpoint of instruction, schemata provide the organizational mechanisms by which new information is related to previous learning, serving as the framework for understanding new material. The schemata brought to the learning situation comprise what we have referred to as initial status or capability (skills and abilities taken together), and determine, as Horton and Mills (1984) describe, ... "what will be learned and remembered and how well." The collective schemata represent the "prior experience" variable noted as critical in understanding the characteristics of acquisition curves. Recall of information or retention of a skill is a function of the degree of development or strength of the schema. (Farr (1986)

further develops the theoretical relationships of schema to retention.) Well-developed schemata also allow for the prediction of future instances or for the detection and filling in of missing information required for new information to be coherent with the existing schema structure.

There is a strong emphasis in schema theory on the role of verbal mediation in skill development. In Anderson's (1982) descriptions of acquisition, learning starts with verbal (declarative) knowledge as the initial structure for generation of behavior. Throughout the schemata-oriented description of learning, verbal coding serves as the initial basis for formation of procedures; the propositional networks for representation of accumulated knowledge are semantic in nature (Neves & Anderson, 1981; Norman, Gentner & Stevens, 1976).

Despite the apparent emphasis in schemata terminology and description on verbal and procedural skills and the role of "knowledge," the theory has been widely used in the study of motor learning. Broadly defined, knowledge in schema development includes the relationships between system inputs and system outputs and makes provision for the development of motor control within the same framework as that used for the acquisition of verbal knowledge. As noted by Gopher, Koenig, Karis and Donchin (1984), "Current models of complex psychomotor skills conceptualize the generation of skilled movements as being governed by high level motor programs or schemas in long-term memory developed with practice."

Schmidt (1975; 1982) and Shapiro and Schmidt (1982) treat motor control as a prime application of schema theory. In their development, schema formation and the tuning and restructuring of schemata in motor skills lead to the formation of "motor programs." These are structures in memory which provide for the "automatic" control of very long response strings with a minimum of voluntary control. They are self sustaining, and eventually,

with extensive practice, bypass the response-feedback loop, relying on observation of output effects for error detection. If performance is perturbed in some way, the motor program initiates correction automatically as part of the control process without voluntary intervention. Schema which guide the responses of the motor program are referred to as "recall" schema; those which evaluate outcomes and consequences and make inferences about expectations are "recognition" schema.

The motor program conceptualization derived from schema theory shares a considerable theoretical basis with the closed-loop theory of Adams (1971). Both treat motor learning as a process of perceptual control and feedback. Adams' conception of the "perceptual trace" is in essence what is learned in the formation of the motor program. Both theories view the initial development of control processes as a verbally mediated activity, and the refinement of processes as an improving ability to detect errors through the learning of discriminations which allow for progressively better recognition of "good" performance. This emphasis on "cognitive" activity in the "modern" analysis of perceptual-motor learning is thus external to, but consistent with, schema theory explanations (Adams, 1985). The formation of highly refined discriminations among task and environment features ("percepts") as filters for input processing and as guides to motor learning has, over time, gained impetus in the study of skill transfer (Lintern, 1985; Vanderplas, 1958). Such percept formation may be a major explanatory concept in both the improved use of simulators for training and in reduction of aversive reactions to simulators through countertraining (Berbaum, Kennedy, Welch & Brannan, 1985).

The mechanisms posited by schema theory for the phenomenon of automatic behavior, though cast in terms of restructuring and tuning of schemata, predict a course of automaticity development that is practically indistinguishable empirically from that

based on the capacity and attentional mechanisms previously discussed. Transfer is important in both approaches, both in itself and as a test for evaluating the predictions from theory. For practical purposes, predictions about transfer would be virtually the same. Transfer is assisted by continued learning to some point in practice; extent of transfer (positive or negative) is a function of similarity between the cues, filters and response relationships of the two tasks. For dissimilar task structures, negative transfer is induced and the amount of negative transfer is related to the degree of practice (unless there are deliberate manipulations of the training task environment to avoid it).

While the theoretical richness of the two approaches would likely yield differences in the prediction of some details of the automaticity and transfer phenomena, the two theoretical developments, taken together, suggest the robustness of the automaticity concept to alternative explanations. We have noted previously the implications of automaticity for exercising judgment about the points for termination of training segments. That conclusion is only marginally affected by differences between theoretical descriptions.

CONTEXTUAL INTERFERENCE

In our previous discussions of focus and emphasis in the analysis of acquisition, we noted the importance of retention as the key variable in examining alternative approaches to training, and described some conditions under which poorer training performance leads to better retention and transfer. This phenomenon has been observed in a variety of learning contexts, e. g., Battig (1966, 1972) and Schapiro & Schmidt's (1982) discussions of the effects of task variation.

Shea and Morgan (1979) presented the term "contextual interference" (CI) to describe the process by which

characteristics of a task create a learning environment in which performance is depressed but learning is apparently enhanced. They deliberately induced interference in acquisition of a basic task by random variation in task conditions during training, reporting materially improved transfer and retention as a result. Lee and Magill (1983) used a similar paradigm, with more frequent systematic switching among various conditions of task properties, and found similar outcomes. Both attributed the improved learning to the increase in cognitive effort (attention) required, i. e., subjects worked "harder" during acquisition because the learning task was more difficult. The effects of variation on transfer may be to a significant extent dependent on the task. Wrisberg and Winter (1983), for example, varied targets on a motor skill task and found no generalizable effect on transfer performance.

In terms of schema theory, the strength or "richness" of the schemata formed under CI would be greater than for one formed under less task variability (Shapiro & Schmidt, 1982). Variable practice enhances schema development by offering the opportunity for the trainee to observe a wider range of associations among task features, responses and consequences of actions. It produces a more generalizable schema that requires less restructuring when the trainee is presented with novel but related task conditions.

Schema explanations (and those of Shea & Morgan (1979) and Lee & Magill (1983)) suggest the variability of task conditions as the basis for the CI effects. The contextual interference phenomenon would appear to be generalizable to a broader range of learning conditions in which the unifying characteristic is task difficulty. While task variability clearly increases difficulty, task difficulty (and effects which show apparent similarity to those of CI) can be produced by a variety of other factors. A task can be hard to learn because it requires unusual stimulus/response relationships (control/display

reversals) or because the initial frequency of successful responses is low and rate of reinforcement is correspondingly depressed. Such tasks demand the generation and test of a large number of performance strategies with apparently similar outcomes, in effect producing a sort of "internal" interference without the external manipulation of difficulty by task variation.

In an experiment by S. A. Jones and Kuntz (1985), subjects traced a star in a computerized version of the mirror-drawing task. The group trained under the "normal" task condition (a mirror image or left-right reversal) had much superior training performance but greater problems in transfer to a "reversal/tilt" condition than those trained under "reversal" (left-right, up-down) or "tilt" (mirror image with 45-degree rotation) conditions. The two more difficult tasks, while performed much more poorly during training, transferred or "generalized" more readily to the more demanding transfer condition. This enhancement of transfer by manipulation of difficulty may not hold for extreme cases. Data from Shephard and Lewis (1950) and Fitts and Seeger (1953) suggest that subjects trained under control/display reversal conditions may "never" (at least in any practical period) attain the performance under "normal" control/display relationships of those trained under normal conditions from the start.

The implications of CI for training are that variation of task conditions, manipulation of task difficulty, or other means of inducing extra effort by the trainee are likely to be beneficial for retention and transfer. Literature reviewed by Cormier (1984) suggests that "moderate" variation in the training task promotes transfer performance. Lintern (1985b) suggests that training approaches (particularly in simulators) should capitalize on CI effects by the deliberate addition of relevant cues to the task environment. He suggests that the positive effects of CI can be obtained without necessarily

sacrificing training performance, and proposes a "contextual facilitation" approach based on a more "cue-rich" training situation, in particular more visual information. The trainee, dealing with these additional relevant cues, would develop the more generalizable schema that lead to greater transfer. Cormier's (1984) review supports this increased emphasis on cues in transfer. He suggests that simulator "fidelity" issues should address the nature of cue complexes which evoke common responses in transfer and training tasks, rather than the common elements of activities involved in task execution.

Findings on CI taken together also suggest that the negative transfer known to occur when overlearning is continued for too long a period may be to some degree avoidable. It is believed that such transfer is due to the development of task-specific patterns which are progressively less generalizable with practice. The manipulation of task conditions to increase the range of instances considered in schema development may produce from the outset a broader definition of the "task" and result in more, rather than less, generalizability as practice refines the efficiency of discriminations among task properties. This is likely to increase training time, but may in the long term be more resource efficient.

KNOWLEDGE OF RESULTS AND FEEDBACK

As with most of the acquisition issues addressed above, the literature on knowledge of results (KR) and feedback is extensive. As Salmoni, Schmidt and Walter (1984) point out, recent data on KR suggest that the bulk of this work has dealt with tasks (predominantly motor) too simple to be directly generalizable to the complex real-world tasks with which modern analyses of acquisition are concerned. As they note, however, and as Schendel, Shields and Katz (1978) reinforce in their review of retention, KR is second only to practice as the critical variable in skill acquisition. Progress in early

learning stages cannot occur without KR, since "good" and "poor" performances cannot be differentiated by the trainee. (In Adams' (1971) terms, the "perceptual trace" has not yet formed). Only when skill has developed to a point at which autonomous good/poor discriminations have emerged can learning proceed without KR.

Salmoni et al. (1984) make distinctions between feedback, KR, and KP (knowledge of performance) similar to those made by others between intrinsic, extrinsic and augmented feedback. They consider feedback as the direct return of information from a sense organ about the nature of an initiated response, KR as response-produced feedback supplemented by information about the outcome of the response, and KP as KR further augmented by information on the actual response relative to the desired or correct response. KR in their descriptions combines the "intrinsic" and "extrinsic" feedback used by some other writers, and their KP is similar to notions of "augmented feedback" elsewhere in the literature.

The findings of Salmoni et al., although predominantly concerned with motor skill acquisition, generalize well to the multi-component tasks typical of military training, as do the portions of Schendel et al. (1978) and Cormier (1984) which deal with KR issues. The latter two reviews are specifically keyed to the military training context. The three summaries taken together subsume most previous reviews of the topic (e.g., Adams, 1968; Adams, 1971; Bilodeau & Bilodeau, 1961; Kulhavy, 1977), and we will deal primarily with their comments and indications.

A key point of Salmoni et al., one that has been repeatedly noted in our previous discussions, is that it is necessary to separate training performance from learning to evaluate the effects of KR. Much of the literature on KR has found an impact of KR on performance as a result of the manipulation of task

conditions. As both Salmoni et al. and Cormier emphasize, transfer is the proper test of KR effects, and changes in training performance, which may be transient, must be separated from the relatively permanent effects of KR due to enhanced learning. As noted above, KR in early practice is essential to learning, but it can be overemphasized as a agent in causing learning to occur. Salmoni et al. point out that KR does not in and of itself cause the generation of correct responses, and that the major function of KR throughout most of the learning process is to help the learner to find the proper responses on his own using other exploratory mechanisms. In that connection, Cormier discusses issues of guidance vs. discovery training, which relate to the extent of assistance provided the trainee in developing a set of correct response mechanisms. For more difficult or more complex tasks, guidance appears to be essential if responses are to be acquired within practical time frames. If, however, only one correct way is provided (as in much military training), the generalizability of skill is poor (an incomplete schema), and transfer and retention are adversely affected compared to learning in which a variety of correct and incorrect alternatives are explored.

Such supplements in training to the normal conditions of KR and feedback must be used with caution if transient performance effects are to be avoided. In general, the literature suggests that "augmented feedback" (KP) primarily influences task performance rather than learning (Cormier, 1984; Kinkade, 1963), and performance increments tend to vanish when augmentation is removed. The major exceptions to that general finding are found in work reported by Lintern and Roscoe (1980) and Lintern (1985a). Both deal extensively with the augmentation of visual cues used in simulator training, and report a moderate transfer effect. Lintern (1985a) develops a perceptual-learning explanation for the enhanced transfer which is similar to his argument (Lintern, 1985b) for contextual facilitation through greater attention to visual cues, and which is consistent with

Cormier's (1984) suggestion that cue commonality is the appropriate basis for determining simulator fidelity requirements. In the relatively cue-poor environment of visual simulators, cue augmentation or enhancement provides a closer approximation to the real task of system control, and is likely to produce findings of improved transfer, whereas augmentation for more procedural or rule-oriented tasks would show no long-term effect.

There are suggestions in the literature that the quality and nature of "guidance" (KR and KP) may have as large an impact on complex skill learning as whether or not guidance is provided at all. There are many highly skilled tasks for which the learner would never reach the advanced stages of performance on his own, no matter how long a period of "discovery" is allowed. Clearly, some techniques must be taught through demonstration, "coaching" or the like. Fischer, Brown and Burton (1978) describe tasks of this nature, for which the amount of simple practice is less important than the conditions under which practice occurs, and simple feedback and KR are insufficient without detailed guidance on how to detect and correct errors. They use skiing as an example of such a task (flying an aircraft involves similar considerations), in which response patterns that produce acceptable performance in intermediate stages are incomplete and must be systematically adjusted or corrected by intensive application of (sometimes one-on-one) guidance principles. They suggest the decomposition of the total task into a series of "microworlds", each more complex than the preceding one, with distinct intermediate goals and a learning environment appropriate to the complexity of the microworld. For their skiing example, progressive microworlds might be "short skis, smooth terrain," followed by "longer skis, smooth terrain," and eventually "long skis, complex slopes."

The notion of decomposing a task, not into its separate skills or into a part-task sequence, but into simplified whole

tasks in simplified environments, has a considerable application for many kinds of complex military tasks (including flying, vehicle control and potentially a variety of information integration tasks). Unlike conventional guidance training, it offers the opportunity for the learner to explore a variety of correct and incorrect alternatives, thus strengthening the generalizability of the "schema" or "program," without the risk of fixating on less than optimal intermediate patterns that are insufficient for the task in its full complexity.

The structure that has evolved over time for military flight instruction has many of these properties. The use of simpler, more "forgiving" aircraft for early instruction, the practice of carrier landings on airfields prior to carrier qualification, the intensive interaction with instructors for "error correction" in the early stages of each new skill, all are more consistent with the "microworld" concept of task decomposition than with the conventional part-task segmentation.

Two other views of the aviation training process are consistent with this "simplification of environments" concept. Eddowes (1974) considers the learning of flight skills as a "spiral expanding cognitive process" rather than the "linear perceptual-motor refinement" process with which it is approached from a motor skill orientation. His description of refinement of cognitive discriminations and the "spiraling" difficulty of successive task segments is a clear statement of the "simplified microworlds" model. Martin (1984) maintains that the fundamental weakness in aviation training is the failure to provide sufficient practice on the fundamental skills early in training. Trainees are moved along to later segments as soon as they show minimum mastery of the "simpler" skills which will form the underpinnings of more complex ones, and the fundamentals never become "fixed." In her view, early attention to error correction focuses on simpler skills, and is crucial to later proficiency.

PART-TASK AND TASK SIMPLIFICATION APPROACHES

Describing the nature of acquisition under part-task approaches is complicated by the vague usage of the terms "part-task" and "task." Task is used to describe a broad class of activities ranging from a procedure with only a few steps to a complete "job" composed of numerous sequences and "subtasks." The term "part-task" has been applied in at least three distinct ways: a) To major procedural segments of a larger task (e. g., the final segment in approach and landing), b) to a special condition of a complete activity (night carrier approach and landing) and c) to the practice of a single component skill of a complete segment (missile envelope recognition in air combat training). The implications and efficacy of part-task training are different for these three usages. Since we are concerned primarily with real-world task representations, a "task" for our purposes is a significant collection of activities which requires some degree of focused effort to learn, and involves some natural sequence of events or segments within the task structure. For precision of terminology, some distinction among the various usages of part-task is desirable. The following breakout seems consistent with the literature previously cited. It is similar to the categories of part-task manipulation used by Wightman (1983) and Wightman and Lintern (1985), but differs in its view of "task" as a larger unit than the usual experimental tasks in part-task studies, which are heavily motor or control in nature.

a) Task simplification -- decomposition and redesign of the entire learning task into a sequence of less complex "whole tasks" with defined but limited objectives (as in the microworlds of Fischer, Brown & Burton, 1978). Each segment requires all or nearly all of the component skills or informational concepts of the complete task, but in lesser amounts early in the sequence, systematically increasing in difficulty as practice continues (Eddowes' (1974) "cognitive

spiral"). This is a somewhat different concept from the usual use of "simplification," which involves adjusting difficulty of task parameters within a whole-task setting.

b) Skill decomposition -- breaking down a task or segment into the component skills or concepts required for its execution (motor, procedural, planning, etc.) and providing specialized practice or pre-practice on individual skills prior to or supplemental to whole-task practice. This is related to concepts of "fractionation," in which independent practice is provided on major control parameters prior to practice with their joint effects, but extends to knowledge or informational components as well.

c) Segment decomposition -- identification and isolation of a complete segment of the total task, "lifted" essentially intact from its place in the normal sequence of segments that make up a task. Additional practice is provided on the segment before integration of segments in whole-task practice. This is the most common "part-task" approach, and has had the widest use and evaluation in military training, particularly in aviation. This is essentially the temporal decomposition aspect of the "segmentation" approach (Wightman, 1983).

Part-task approaches have been shown to be effective in a variety of learning situations. Predominant use historically has been in the training of motor skills, although there has been progressively more use in recent years in the training of procedural tasks, particularly in maintenance (Orlansky & String, 1981; Wisher, 1985), and in specialized skill trainers for very broad and difficult "components" of military jobs, i.e., Ricard and McWilliams's (1985) "expert systems" trainer for training antisubmarine warfare tactics. A number of reviews and analyses have documented the effectiveness and appropriate uses of part-task training (Adams & Hufford, 1962; Naylor, 1962; Orlansky & String, 1981; Schendel, Shields & Katz, 1978;

Wheaton, Rose, Fingerman, Korotkin and Holding, 1976; Wightman, 1983; Wightman & Lintern, 1985).

In particular, task simplification approaches are likely to be beneficial for tasks of the type described by Fischer, Brown and Burton (1978), in which skill requirements are very high, appropriate strategies are unlikely to be developed by unguided exploration, and risks are high of reinforcement of incomplete or partial strategies that are insufficient but hard to eradicate in later stages. Skill decomposition is appropriate for tasks in which mastery of one particular skill or concept is both more difficult and more critical for successful performance and can be practiced independently of other components. Similarly, segment decomposition will improve performance and/or efficiency when a limited number of critical segments or sequences are materially harder to perform proficiently. Two factors bring about this improvement. If the segment is terminal in a conventional training sequence, it must be performed in early trials from a starting point based on the cumulative errors created in previous segments. If it is more difficult, regardless of its position in a sequence, it will be the last segment mastered; training is likely to be terminated before "mastery" is complete, producing poor transfer and retention, and other segments are likely to have been heavily overlearned in conventional training, wasting much of the time spent in practice.

Task or job decomposition can be carried out with a variety of approaches and methodologies, depending on the nature of analysis or decomposition desired. Techniques such as those described by Miller (1962), McCormick & Jeanneret (1984), Sparrow, Patrick, Spurgeon and Barwell (1982), and Mane, Colès, Wickens and Donchin (1983) may be applicable in deriving an appropriate breakdown of tasks into component skills and segments. The technique used by Mane et al. (1983) is particularly well-suited to part-task training analysis; Mane (1984) gives an example of its use in skill acquisition.

PREDICTION OF TIME TO TRAIN

There are, without question, massive individual differences in the time required to learn a given skill or body of material. We discussed earlier the distinctions between time required to learn (TTL), and time spent learning (TSL) and the advocacy of the relationship between these quantities for an individual as a key indicant of how well learning would be retained (Gettinger & White, 1979; Gettinger, 1984). The prediction of how long training should take is equivalent to estimating the point at which TSL is equal to the average of individual TTL values. It is well-documented that learning performance interacts with method of instruction so that some individuals learn better with one method or treatment than with another (Cronbach, 1957; Cronbach & Snow, 1976; Frederickson, 1979). Although these interactions are most readily observed in individuals of different ability levels, they can occur also from subtle differences in the mix of abilities brought to the learning situation. As Cronbach (1975) has observed, the effects are most often seen as complex interactions instead of simple tendencies, and are difficult to handle in practical training and education situations.

If relative time required to learn could be estimated from ability measures prior to training, "appropriate" training methods (however determined) could be prescribed. Only limited work has been done on formal quantitative methods for assignment to a training group based on measured abilities. Williges and Williges (1980) report on a series of experiments as part of a larger program for individualizing motor skills training. Savage, Williges and Williges (1982) document a specific experiment from that series and discuss the few studies on similar topics. Williges and Williges (1980) introduce the concept of "macroadaptation." Just as adaptive training attempts to compensate for large individual differences in time required to master a skill, macroadaptation looks for a limited

set of learning styles or approaches as a function of ability so that a limited number of alternative training strategies could be used, matched to learning style.

In the studies reported by Williges and Williges (1980) and Savage et al. (1982), regression models were used as decision rules for assignment to a training strategy. Two instructional methods were applied, either fixed difficulty or adaptive learning of a tracking task. The prediction battery consisted of six paper and pencil tests, heavily weighted toward spatial and figure-ground abilities. In a pilot study, scores were related to performance obtained under the two training conditions and assignment formulae were developed. Individuals in a new group were assigned to either matched (condition with highest performance prediction), mismatched, or random groups.

Several novel outcomes were found. a) As hypothesized, time required to achieve criterion was much less (50 percent) for matched groups than for random, with an even greater differential for mismatched. b) Surprisingly, students in the matched and random groups did significantly less well in transfer than the mismatched. This is highly consistent with our previous discussions of contextual interference and the role of task difficulty as a factor in enhancing learning, as distinct from the effects of simple variation in task conditions. c) The equations used for assignment did not work effectively for a different population. When applied to students of a less variable academic background (military academy), equations dropped substantially and the expected effects did not appear. d) There were, for both populations, major differences between males and females in predictability and in the variables effective for prediction.

These experiments taken together suggest that given a relatively simple task, sufficient understanding of the task and its constituent abilities, and care in developing of assignment

strategies, it is possible to improve the efficiency of training through clustering of individuals into different training method groups. They also suggest, however, that transfer can be adversely affected by doing so, and that the formula for prediction of time to train may be both highly task-specific and highly sensitive to variations in characteristics of the trainee population.

The attempt in these studies to find subsets of the trainee population which would benefit from different strategies is similar to one of the main initial objectives of the present analysis -- to find ways of clustering individual learning curves into families of "learning styles", which might be relevant to determining the time course of training. The findings of Williges and Williges and Savage et al. are consistent with those of the present analysis -- while it is clear that such families of curves exist, their parameters and expected shape are likely task-specific and would require a considerable amount of historical data to implement with confidence in an applied training setting. They also require good predictors, probably tailored to a task situation, and data routinely available on trainees (e. g., AFQT/ASVAB) are not likely to give acceptable differentiation. If it is desired to capitalize on individual differences in rate and patterns of learning, the most feasible route is to track individuals during training, using immediate past performance as predictors to estimate the future course of performance, and use the intervention strategies we have noted for "mid-course" corrections in training approach. Spears (1985) notes a similar potential use of in-training "curve" data.

RETENTION

We have not, in discussions of acquisition, attempted to summarize in any depth the highly complex literature on retention and transfer. We have, however, used a number of

references as background for analyses and conclusions and for selection of emphasis in acquisition studies. A representative sampling of the major retention and transfer studies on which we have drawn is included in the reference section. In addition to the major efforts by Cormier (1984), Schendel et al. (1978), Hagman and Rose (1983), Farr (1986), and other reviews already cited, source materials also included the 22 retention and transfer articles identified in the reference section with an asterisk.

ISSUES IN TASK AND SKILL CLASSIFICATION

Applied training can be viewed as a transformation of some "raw material" (the entrant population) into an "output" (graduates) that has the capability to do one or more of a variety of jobs or to benefit from further "transformation" through additional training. The process requires a) a way of determining what people need to do (requirements of a job or training task), b) a way of categorizing the skills and knowledge that people can have at entry or acquire through training, c) a means of mapping initial status into completion status to derive a set of training prescriptions which change raw material into output, and d) a test or evaluation procedure to determine if the output meets the "design specifications." Application to training of the considerable body of "how to instruct" information previously discussed has been hindered by the lack of effective mechanisms for accomplishing these objectives, in particular by the absence of consensually acceptable schemes for categorizing task requirements and for classifying skills and knowledges.

The "taxonomy" problem in training and its impact has been repeatedly articulated. Ferguson (1956) recognized the lack of systematization in learning and defined the problems of not having taxonomic descriptions of differences among learning tasks. Miller (1962, 1975) presented an outline for task

analytic description and emphasized the need for a common framework for both training and task requirements. Glaser and Resnick (1972) reinforced the need for identification of the basic kinds of learning processes as a means of guiding research. They discuss the basic distinction between the two types of classifications required, one which organizes the "learning conditions" or task requirements (the desired outcomes, as in Gagne's (1977) structure), and one which organizes the inferred learning processes presumed to underlie task performance. Glaser (1976) refers further to needs for a "linking science" to marry concerns with individual differences, lawful knowledge about behavior, and issues of categorization into a science of instructional design.

Categories of Classification Systems

A number of writers have pointed out the criticality of taxonomies and classifications to making progress in theory and in applied arenas. In addition to those already cited, Fleishman (1982) and Fleishman and Quaintance (1984) systematically link the general availability of task and skill/ability structures to the informed guidance of future research and to the effective use of existing data. The Fleishman and Quaintance volume brings together and relates virtually all attempts to develop categories for describing tasks and processes. They suggest three ways in which tasks can be classified or described -- a) in terms of the behaviors involved in task execution (Behavior Description), e. g., Berliner, Angell and Shearer (1964); b) in terms of behaviors required to perform a task (Behavior Requirements), e. g., Gagne and Briggs (1979); and c) in terms of abilities required to perform (Abilities and Task Characteristics), e. g., McCormick and Jeanneret (1984). They map some 35 previous classificatory systems into these categories.

Categories of Taxonomies

Fleishman and Quaintance make a distinction between classification schemes and taxonomies, most notably in the requirement for taxonomies to be hierarchically arranged. They provide five basic structures for approach to taxonomy development, which are in part generalizations and extensions of their breakout of classification systems. These general categories are:

a) Criterion Measures -- tasks are described in terms of type of output or performance variable.

b) Information-Theoretic -- tasks are defined in terms of information transferred between information source and receiver.

c) Task Strategies -- tasks are described in terms of sequences of transaction events between operator and environment.

d) Ability Requirements -- tasks are defined in terms of the abilities required to perform them.

e) Task Characteristics -- tasks are described in terms of stimulus-response, goal relationships and other descriptions independent of human characteristics.

One of the most complete systems for task characterization is the Berliner Behavior Classification Method (Berliner et al, 1964). It was developed primarily for description of operator performance, and was applied in that context by Christensen and Mills (1967) to catalog the specific activities required for tasks in operation of a complex weapon systems. The method breaks operator behaviors into the type of Process involved (Perceptual, Mediatlional, Communication and Motor), the Activities (Search for, Identify, Process, etc.) and the Specific Behaviors (Inspect, Monitor, Compare, Remember, etc.).

Christensen (1982) also used the the Berliner classification to link a set of desired behaviors on a (highly procedural) task to the formal Instructional Systems Development process.

The Rasmussen (1979) paradigm previously presented (skill-based, rule-based and knowledge-based tasks), while not a formal classification system in the sense of those above, involves elements of several of the Fleishman and Quaintance categories. It deals with both the behaviors involved in task execution and the general nature of the skills required. In its hierarchy of least to most complex behaviors, it defines the gap between entrant status and desired output capability which must be narrowed by the training system, and has potential for linking task content (in behavioral terms) to the type of training required for proficiency.

A full review of taxonomies and classificatory efforts is well beyond the scope of this analysis. We believe, however, that serious attention should be given in military training research to better description of training-related abilities (and skills) and of task requirements, to allow a more orderly mapping of research findings into the planning of training. We believe also that a better understanding of the processes underlying behavior is critical to understanding the driving factors in the shape and form of acquisition curves. In the mathematical developments of Anderson (1982) and Newell and Rosenbloom (1981), the way processes bring about learning and the nature of their interactions with one another determine the basic functional form of skill acquisition curves. In their analysis, verification of curve shape predictions is in part a verification of those processes and mechanisms, which in turn are a key element in the cognitive-theoretic prescriptions for achieving a desired course of learning.

GENERALIZATIONS AND DISCUSSION

OVERVIEW

We have in preceding sections presented both theoretical and data-interpretive evidence relevant to determining and manipulating the time course of segments and phases of military training. We have examined families of curves for data that describe the process of skill acquisition, compared those families of curves to empirical findings, identified factors that influence curve shape (i. e., rate of learning), and attempted to abstract from these analyses some regularities that might be useful in bringing about more efficient and/or more effective approaches to training. We have also discussed the characteristics of the military training environment that can influence skill acquisition and some of the constraints under which that environment must operate. The predominant subject matter of the present analysis involves training approaches which concern the time course and schedules of training in military settings.

In this section we bring together some generalizations derived from our review and analysis of acquisition data and theory. Generalizations are of two major but distinct classes. The first are generalizations drawn directly from considerations of the technical data and the theories to which they are relevant. These findings and insights are documented in a number of earlier sections. In most cases, they are restated here in an abbreviated format without further presentation of evidence, and are referred to as "generalizations from the data."

A second class of general statements is aimed at the functions and uses of learning data in the military training environment. These represent both a synthesis of evidence previously cited and an extrapolation of that data to make inferences about a) what kinds of data are needed to improve

military training, b) where and how could these data be obtained, and c) what applications of the data are likely to be most appropriate and effective. These are "generalizations about the data" on learning and acquisition, and involve a series of judgments about the current uses of learning data and principles in military training compared to what it is possible to do and what "ought to be" done to obtain more effective training. Such generalizations are thus based not just on what evidence and data are available in the literature, but also on what is not there, and perhaps should be.

GENERALIZATIONS FROM THE DATA

Interpretation and Uses of Learning Curves

o There are, throughout the literature, some striking regularities in the general form and appearance of learning and acquisition curves: The negatively accelerated curve is not only "typical" of group performance (and many individual performances), but is likely to be found in nearly all learning of the "practical" tasks characteristic of military training situations. Recall the representative curves given in Figure 1. The general shape of these curves is consistent with all major theoretical explanations of how skill acquisition proceeds. This finding is sufficiently pronounced that, when curves of other shapes are encountered, some examination of task composition and segmentation, the method of instruction and the difficulty of the task relative to the trainees' abilities and experience may be warranted.

o While the negatively accelerated shape is the overwhelmingly common form of the practice curve, parameters of the curves tend to be task-and-situation dependent, and are not readily generalizable to other situations. Further, there does not appear to be any one single "best" or universal mathematical function for describing such growth of skill with practice. The

particular curve family providing the best fit to a given set of data is likely to vary as a function of (among other factors) the task, its components and its level of difficulty, the characteristics of the people performing the task, the length of practice, the way in which performance is measured, and the degree of innovation involved in the training method used. The power function (and hyperbolic) equations come closest to a "universal fit," and offer curves with interpretations well-anchored in theory, but the power function cannot provide satisfactory fits to a portion of the empirical acquisition curves. Reasons for the effectiveness or ineffectiveness of the various functions in fitting empirical data are not well understood. There have been too few systematic comparisons of alternative functions to determine which curve families are likely to be most appropriate for a particular set of tasks and task conditions.

- o It is often difficult to interpret acquisition or practice data in the literature because key information, such as subjects' abilities and experience, is often omitted. Inconsistencies between learning patterns and curve parameters on what are apparently the same or similar tasks cannot be resolved without this information.

- o Plateaus, i.e., periods of practice during which no perceptible performance changes are seen, are common in learning real tasks, and are an "interruption" to the negatively accelerated growth curve. While they complicate the mathematical description of practice, their existence has important theoretical implications for understanding the processes of skill acquisition and integration and for predicting the time course of training.

- o The degree to which individual learning curves are representative of group learning patterns (or group learning is representative of individual learning) is highly variable. The

departures of individuals from group curves increase as the group becomes more heterogeneous in ability, as the task becomes more complex (multi-component), and as the task structure allows for successful performance through a variety of different strategies.

More Emphasis on Acquisition than Retention

In studies of how people acquire skills and knowledge, there has been much greater emphasis on the acquisition or training aspects than on the long-term retention of skills and knowledge and their use in other settings. This complicates the use of learning data for training improvement since some instructional strategies can improve apparent learning without impacting retention or transfer. Recourse to theory is ultimately required to resolve these complications. Work in schema theory and on concepts such as "contextual interference" has begun to clarify distinctions between performance and learning.

Selection of Termination Points for Training

o The point in acquisition of a skill at which practice should be terminated is a tradeoff between the objectives of retention and those of transfer, modified by considerations of the cost of continued practice. Retention improves with practice almost without limit, although gains in retention systematically decrease as practice continues. "Overlearning" also improves transfer, but only to the point at which skill generalization is reduced by the development of task-specific "automatic" patterns of responding. Research evidence suggests that it is ultimately necessary to track individual learning curves on specific tasks in order to identify appropriate termination points. The choice of termination point is critical for experimental studies because different points can produce changes in outcomes with distinctly variant theoretical explanations. Choosing the "right" point is important in

training programs because of steadily decreasing return-on-investment after performance becomes stabilized.

o Better retention (and probably better transfer) is obtained if several successful performances of a task are required of a trainee prior to termination of a segment. One successful demonstration or one score above the qualification level apparently leads to termination of training when performance is still relatively unstable. Redefinition of mastery to include some trials past initial demonstration of proficiency also avoids the risks of achieving mastery performance by random variations around true performance. The number of trials of overlearning appropriate to a particular setting should be empirically determined, but recent work suggests that three or four additional trials may be sufficient for most moderately complex tasks.

o Increasing the opportunity of the trainee for successful repetitions also appears to increase confidence in behavior, to reduce non-productive anxiety about capabilities, and to improve resistance of performance to disruption by changes in task conditions. This is important in operational performance of tasks, since the modern battlefield may involve a variety of stressors that act to increase task difficulty and decrease performance reserve (e. g., chemical-biological warfare protection, sustained operations, etc.).

Procedural Tasks

o Procedural tasks present special complications for training. Such tasks are quickly forgotten; the decay of learning is rapid ("days or weeks" is usual), and is approximately a linear function of the number of procedural steps. Overlearning is virtually essential but insufficient for acceptable retention of proficiency. Because retention is heavily dependent on remembering cues and sequencing, there are

definite practical limits on the extent to which adjustments or compensation made solely within the training process can improve retention. Provision must be made elsewhere in the system for routine periodic refresher or proficiency training, particularly for procedural tasks which are job-critical but infrequently performed.

- o Some modifications to training may improve learning and retention of procedural tasks. There are indications that practice in learning procedures should be distributed rather than intensively massed, and that learning benefits from frequent interspersed test trials. Further, much current military training gives procedural instruction without an accompanying orientation to how the procedural steps fit into task execution as a whole. Retention of cues which trigger procedures seems to be improved by appreciation of context for procedural steps, and by some variation in conditions to allow exploration of alternative approaches; the generalization of cues is apparently better under the richer "schema" formed under those conditions.

Such enhancement is related to that expected from an increased use of teaching procedures using "functional" explanations of a system's operation. Functional explanations teach procedures by stressing, for example, the understanding of the role of a piece of equipment in the system and how it relates to other system components in addition to explaining the parts of the equipment and how they work. These are believed to improve learning and thus retention by increasing the richness of the retrieval cues that can guide procedures in a manner similar to that above.

Impact and Management of Individual Differences

- o Individual differences in rate and patterns of learning are endemic to all classes of skill and knowledge acquisition.

The effectiveness of strategies for controlling information presentation and practice is heavily dependent on an individual's ability level and prior experience. Traditional instruction uses a fixed time-to-learn and allows learning or achievement to vary. Other approaches vary the time, pacing and/or scheduling of practice as a function of individual differences. These "tailored" approaches such as mastery learning (criterion-referenced training) thus fix achievement level and vary time of instruction. As such, they are more "effective" and economical, but are difficult to implement in traditional fixed-time settings.

o Some limited evidence suggests that it is possible to manage individual differences by predicting how long people will take to learn a skill and to use that forecast for assignment of individuals to one of several training methods. While this approach reduces training time, it appears to have an adverse impact on transfer of the skill for at least some trainees. Further, the parameters of the assignment equations are likely to be highly population-specific.

Stages of Skill Acquisition

The literature strongly supports the presence of a series of stages within the skill acquisition process; during any given stage, both the observable activities of the learner and the internal processes are qualitatively different from those occurring in other stages. Although there are varying terminologies for different theoretical models, there is evidence that any one of several three-stage descriptions provides a satisfactory general representation of the time course of acquisition.

Simplified Part-Task Approaches

The learning of highly complex skills appears to be aided by a variety of part-task training approaches, in particular, those

which decompose the total task into a series of "simplified whole" tasks. These simplified tasks and the environment in which they are performed become progressively more complex as fundamental skills are mastered. Such an approach is consistent with and supported by an increasing body of schema theory research and other cognitively-oriented explanations for the skill acquisition process.

Contextual Interference

The presence of a "contextual interference" phenomenon, in which poorer learning performance leads to better retention and transfer, is well supported by data. Most explanations have posited variation in task conditions during learning as the main agent in creating the effect. A more general explanation may be that any task conditions which make the task more "difficult" (i. e., requiring greater effort) can within limits bring about a contextual interference effect.

GENERALIZATIONS ABOUT THE DATA

Descriptive vs. Comparative Data

o With few exceptions, data on acquisition performance for complex tasks are descriptive in nature. Data from most experiments do not provide the comparative data required to make precise judgments about how best to train a particular skill in a specific training setting. This scarcity of comparative analyses hinders informed decision making in training. At present, determination of course and segment length and selection of training method are made by evolutionary changes to existing systems, often without availability of feedback about effectiveness of the current structure compared to other alternatives.

Research efforts of the last decade have shown an encouraging tendency to examine the complex multicomponent tasks

that are representative of those in the real world, and are considerably more generalizable than those of earlier periods. But few if any studies have systematically compared alternative approaches for training the skills and declarative information accumulated by military personnel as they move through the segments and stages of learning how to do things required on the job.

o Comparative analysis of alternative training methods seems to be subject to the "Hawthorne effect." Modifying the method of instruction, course content and materials, or instructor appears to cause the shape of the progress curve to vary, typically toward the direction of higher performance. Throughout the literature, any change in the training or instructional environment intended to lead to "better" performance tended to produce that effect. The improved learning per unit time produced curves with higher initial performance (similar in effect to greater prior experience), a steeper slope (greater learning increment per trial), and, under some conditions, higher terminal performance. In the great majority of comparative curves examined, the "improved" method led to a curve that departed in significant ways from the "typical" acquisition shape. It is difficult not to speculate as to what might happen if the improved method were implemented and became the baseline. Would inherent system forces and the natural pace of learning established by trainees over time cause the new method to shift to the shape and time course previously seen in the baseline?

Theory vs. Applications in Studies and Data

o The experimental literature is heavily oriented toward data intended to assist in the resolution of differences between theoretical explanations. Only limited attention has been given to the specific collection of data with the goal of direct

application and generalization to training. (This is a rapidly improving situation. Much more of the recent literature has focused on practical skill acquisition implications). The educational literature, while replete with prescriptive structures based on well-developed and consistent theory, is much less rich in data; it deals with education and training paradigms generated directly from theory and rarely examined in a comparative framework. This is in part a result of the subject matter; "classroom" education in schools or in military academic instruction is among the most difficult areas to study quantitatively because of the lack of standardization inherent in instructor-centered approaches. Lectures and verbal instruction are an indeterminate mix of general principles, standardized material and "show and tell" examples, varying from one instructor to the next and between the same instructor in different classes. Although training may be effective, comparison to baseline is difficult in such settings.

o Some research on learning and training does not generalize well to the unique environment of military training. We have noted previously the tendency of learning curve parameters to be situation-specific. There are thus useful principles, but insufficient empirical data from laboratory or other controlled studies that apply closely to particular military training settings. In addition, military training diverges in several important ways from conventional education and training and deals with a different set of constraints. These differences and their implications are not always adequately addressed in the research and development base. It is not so much the problem that "more" research is needed (although it likely is) as that the framework in which studies are cast should reflect explicit consideration of military-specific issues.

o In military training, skills and knowledge are accumulated gradually across successive segments and phases; conversely,

most literature deals with the learning of one complete (often small) task practiced repetitively. Examining the effects of practice on acquisition thus needs to be done with an awareness that the term "practice" has a variety of meanings depending on the task being studied. We noted earlier that military tasks often differ from those in the experimental literature on the basis of the period over which a complete set of skills and information are acquired. Military training frequently resembles general education in its accumulation of "competencies" over a series of courses, segments and stages of training, but differs in its emphasis on multicomponent tasks rather than on the largely informational skills involved in education. In its relatively long-term cumulative nature, learning in the military diverges somewhat from both education and from conventional definitions of what constitutes a task. With the exception of basic recruit training, most training in the military involves tasks that are not end goals in themselves, but serve as a base for later additions to skills as task integration occurs in subsequent training or experience.

To generalize the literature on "practice effects" directly to military training involves some risks. Factors involved in learning by repetitive practice on a self-contained task may not be the same as those which govern cumulative acquisition across courses and segments. While we have in previous discussions about practice effects generally ignored these distinctions, we believe that some additional research is warranted into the extent to which the ways of structuring training derived from the "ubiquitous" laws of practice would differ from those which focused on a cumulative vs. a repetitive model of acquisition. Such studies would track progress over longer periods of training (at least several segments) and would explicitly use a broader definition of "task," closer to the meaning of a "job," than that used in the typical study of acquisition.

Constraints on Implementation

The literature provides a substantial body of guidance as to how to improve training. Many principles are well supported and can be safely generalized to military training settings. We have noted in several previous discussions the real constraints imposed on innovation in military training by the need to produce a specific number of graduates in a fixed time frame. Approaches which could perturb the timing and sequencing of training segments particularly complicate the predictability of personnel schedules and availability for military assignment, and thus face significant institutional hurdles. As a result, most general guidance has not been implemented.

Sources and Uses of Historical Data

o To obtain accurate data suitable for estimating training progress of a group or for intervention in the training process, it may be necessary to generate it from within each particular training program. The specific characteristics of the training context in which data are to be used are important. The literature is conclusive that specific parameters (notably rate, initial status or experience, and asymptote) required for estimating a group acquisition curve are highly task- and content-specific. In other words, quantifications of time-to-learn and other key parameters do not appear to be generalizable much outside the conditions under which they were obtained. While overall shape is probably acceptably predictable, progress over time is dependent on context. The effects of many factors that cause shape to vary are reasonably well understood, but successful quantification in the general case is unlikely. Variability and time course of progress for individuals are a function of (among other things) the materials and skills being taught, the methods used in a particular training setting, and the entrant skill level of trainees, all of which are likely to be stable within a segment and variable among segments.

Without measurable estimates of the effects of these major factors, we can only describe progress, not predict. It is, however, feasible and valuable to collect and use data from within a specific course or segment in a variety of ways that improve training. Baseline or "target" curves can be reliably established using historical data. Acceptable variation from baseline can be determined in the same way from examination of the progress at key points of individuals who have been successful or unsuccessful in previous classes. Interim waypoints can be derived which define unsatisfactory progress and trigger detailed review of an individual's progress and problems. These provide the opportunity to intervene through change of instructional method, additional training, or other forms of individualized instruction. Likewise, trainees well ahead of expected progress can be given enriched or diversified instruction which may help retention or transfer. Such mechanisms, at least for low performers, are currently used in many training situations, but so far as we are aware are rarely or never grounded in historical performance or completion data.

o Maintaining and using cumulative historical data for "quality control" purposes offers considerable leverage for improved training effectiveness and cost reduction. We have already noted the problems with projecting progress curves and estimating time to train from the general forms of acquisition functions. Using immediately prior performance for a group or individual to forecast performance in future training takes advantage of the most relevant data available. It further buffers effectively against the "overachiever vs. underprediction" syndrome. Those individuals who might, on the basis of other factors, be considered as candidates for training intervention (low ability, previous problems), but who "outperform" their predictions, will avoid the disruptive effects of special treatment, and the resource investment for unnecessary remedial training is likewise avoided. A similar application of historical data for tracking, intervention, and

student disposition was used successfully in aviation training for a number of years in the Navy's "Secondary Selection" system (Shoenberger, Wherry & Berkshire, 1963), for quality control of pilot and naval flight officer candidates (Lane & Ambler, 1974; Peterson, Booth, Lane & Ambler, 1967) and for assignment to future training of varying skill requirements (Ambler & Lane, 1974).

o A further potential advantage deriving from the availability of progress data within a course or segment is the possibility of tradeoffs between training time and training effectiveness. Cronholm (1985) defines a conceptual model of a training system as a sequence of successive "instruction blocks" (essentially the same as our "segment"), through which students progress enroute to eventual completion of training. Type of instruction may vary across blocks, and the output of the total sequence is an integrated set of skills and supplementary declarative knowledge. Cronholm suggests that the training system may get out of "adjustment" if the time spent in each block is not "optimized" in terms of a given block's contributions to the ultimate level of performance. He presents theoretical mathematical solutions to the division of training time across blocks to maximize ultimate performance or to minimize cost.

In Cronholm's development, it is presumed that the learning curve parameters and the transfer function relating time (or practice) to progress are known. Johnson (1980) presents a different mathematical solution with a similar rationale for terminating training on the basis of expected return on investment from continued practice. Both solutions require an extensive set of data on which to operate, of the type rarely available in training settings. The maintenance of historical data records would materially increase the feasibility of such mathematical estimation of segment length, and could be instrumental in putting a rational basis under many training decisions presently made on intuitive grounds.

o If reliable data are available, three major determinants of learning -- time, practice (amount and scheduling), and method of training -- can be manipulated by training managers to achieve the required level of proficiency. Depending on the constraints imposed on making changes in a particular training segment, it may not always be possible to vary or adjust all the major factors affecting level of learning, despite the improved efficiency that might result. Time may be fixed by the personnel assignment system (and not subject to negotiation); resources may not enable methods of training to be modified. So long as one of the three main factors is free to be varied, a manager can compensate to some extent for the absence of flexibility in the others. More extensive practice can be given in a fixed time, and the degree of distribution of practice can be changed; method of training may be modifiable while scheduling is fixed, and so on.

If none of the above factors are adjustable, training effectiveness can be materially degraded. Most military training is structured around time as a key variable, not performance. Performance requirements and attainment of sufficient proficiency to succeed in later segments are often subjugated to scheduling issues. This has major implications for retention. Spears (1983) comments with some frustration on a procedural task trained on a single day with extensive successive repetitions, without pauses or breaks and with no testing trials. He reports a) failure rates of 85-93% in field trials a few weeks later, and b) a "strong resistance" by the training organization to modifying the approach, since only one day was "available" without rearranging other scheduling within the segment.

o Both group and "individual" curves are required to use progress data effectively. Group progress curves based on historical data serve as the target for assessing overall progress of both the class and the individual. Points on the

"curve" are empirically obtained, and need not correspond to any particular functional form. Individual curves are obtained for each trainee as he/she moves through the segment. These serve for "quality control," detection and diagnosis of problems and prescription of remedial training.

o Although cumulative historical data on training performance has high immediate and potential value, the effectiveness of decision making within a training segment is ultimately dependent on feedback from later periods about the performance of that segment's "product." In most military training settings, the end-of-course status of a trainee cannot readily be translated into meaningful forecasts of retention or performance in the next segment or on the job, often because the system has not collected the necessary long-term data. The generalities about effects on retention available from the literature are helpful as broad guidance, but cannot be substituted for the follow-up of graduates in a systematic program. The "push 'em through" philosophy reflected in the event reported by Spears (1983) may be an exception, but such tendencies are difficult to discourage without clear documentation of the effects and costs in later stages of "passing on the problems."

o Much of the emphasis placed on the importance and uses of progress data resembles the arguments advanced in support of the scheduling and diagnosis/remediation functions of Computer-Managed Instruction (CMI). A fully implemented CMI system would in fact provide the data required for the type of progress tracking defined above. It may be that our conclusions about the need for data are equivalent to a concurrence with Dollard, Dixon and McCann (1980), Swope, Corey, Evans and Morris (1982), and many others that properly designed computer-managed training can be a major improvement over current methods, particularly in its enabling of training intervention while maintaining predictable student flow.

Instructional Systems Development and Classification Systems

o There are provisions in military training for systematic development of delivery systems. These established procedures are based conceptually on comparing the capabilities of people at entry level with the capabilities needed for competent job performance, with the required training derived in some way from the differential between these two levels of skill and knowledge. The formal Instructional Systems Development (ISD) process is an attempt to regularize steps in those procedures for building training packages for new and existing systems. The specification MIL-T-29053B(TD) (Requirements for Training System Development) calls for a series of analytic efforts. In general, it requires a) a training Task Analysis of activities to be performed on the job, identification of job objectives, etc. (task requirements in our terminology); b) a Critical Task Analysis (CTA) determining the most important or "core" tasks required by the job, c) a Skills and Knowledge Analysis (SKA) which translates tasks into the "capabilities" needed to perform the job, and d) for new systems only, an analysis of entry level skills and knowledge possessed by the "average" trainee. The SKA is then used to define delivery systems, equipment and packages and to provide guidance in media selection.

While the ISD process has been instrumental in focusing attention on the need for systematic treatment of training systems, it describes "what" should be done, not "how." Its effectiveness is limited by the unavailability of consensually accepted schemes for describing job requirements and associated skill and knowledge requirements. Job and task analysis and SKA terminology and procedures used are typically idiosyncratic to the analyst.

We discussed earlier the plethora of classification schemes and taxonomies for both task content description and for the

structure of individual capabilities. Much of the research that ought to be used in training systems development is difficult to apply because of the problems in converting one scheme into another. Consistency in classification structures is a two-edged sword. It is likely that any structure that could achieve broad acceptance would be rudimentary and incomplete; the premature fixing of such a system as a "standard" would almost certainly inhibit progress toward a more mature and comprehensive version. On the other hand, regular use of new findings on how to train for very complex systems will likely require some general "notational" structure, a way of organizing tasks and skills sufficiently comprehensive to allow mapping of a variety of different schemes onto a common base.

Training Effectiveness Data

o It is constantly necessary in evaluating training effectiveness from within-segment data to distinguish between learning and performance. Factors which increase or decrease training performance do not always have a concomitant effect on the level of learning or the effectiveness of training. Learning as the ultimate goal of training is appropriately measured in terms of resistance to forgetting or by its contributions to learning a related task. Assessing effectiveness, particularly in military training, requires the tracking of training progress and performance across some minimum number of segments to determine either the satisfactoriness of present methods or the relative goodness of alternative approaches. There are strong indications in the literature that some methods of training that increase difficulty in one of several ways will cause lower performance in initial training, but higher performance later during transfer and retention. Statements about changes in effectiveness based entirely on data from within a course or segment can be misleading unless direct relationships have previously been established between end-of-course status and later performance.

o Training effectiveness is broadly defined and may be reflected in variables other than the retention or transfer of performance. Besides job performance improvements, there are other benefits of training that should be considered in evaluating how well a training program is working. Different levels of learning that result in equivalent performances under normal task conditions can produce significant variation in performance when conditions are changed. As noted above, overlearning and other training enhancements result in increased resistance to disruption and in an improved capability to sustain an acceptable performance under degraded mode conditions. These are important aspects of effectiveness analysis and are more closely related to the implicit goals of training than are the more conventional metrics of in-course performance and retention and transfer obtained under nominal or standard task conditions.

CONCLUSIONS

1. The "typical" curve relating training performance to practice has a characteristic negatively accelerated shape. Curves deviate frequently from that common shape, but usually as the result of one or more well-understood task characteristics or training conditions. The shape is sufficiently regular to form a "baseline" or target curve for training progress; major deviations from the general shape may represent undetected aspects of a training program that produce either efficiencies or inefficiencies.

2. While curve shape (general appearance) can often be anticipated reliably, the time-course over which acquisition runs, and thus the curve parameters, is generally not predictable from prior knowledge of task characteristics. The mathematical description of a task learning curve requires data specific to a task or training segment.

3. There is much useful information in the learning and skill acquisition literature that could be applied to improve military training. The most valuable of these principles concern time-based aspects of training -- sequencing, scheduling, pacing and course length -- rather than training content per se. With a few exceptions, the benefit of these principles has not been realized in military training situations.

4. The main constraints on use of time-based principles for making decisions in military training are:

a. The fixed-time orientation in military training arising from the need for predictability of personnel availability for assignment.

b. Available data from the learning and training literature are not always sufficiently task-specific for generalization to military tasks.

c. Available data tend to deal with the practice of complete tasks, and do not generalize well to the cumulative acquisition of skill across training segments typical of military training structures.

5. Understanding the nature of learning which occurs on a specific training task as practice continues is critically important, both for efficient training and for avoiding undesired negative transfer effects. Determining the point at which "sufficient" training has been given (i. e., course or segment length) is complicated a) by a lack of task-specific retention and transfer data and b) by the difficulty of isolating those task conditions which enhance learning from those which improve training performance without enhancing learning. A number of recent developments in the theory of skill acquisition, such as contextual interference, have potential for resolving the latter difficulty.

RECOMMENDATIONS

OVERVIEW

The following recommendations are restricted to those which have direct applicability to military training. Other recommendations, not directly germane to the main theme of time course estimation and manipulation, are found throughout the paper. A brief rationale is provided for each recommendation.

1. COLLECTION AND USE OF TRAINING DATA

Establish mechanisms in military training for the routine collection, maintenance, analysis and application of progress and performance data. Data should generally be collected and maintained at the segment or course level. Mechanisms can be separate from or supplemental to any existing computer-managed instruction, but they should produce data sufficient for a) defining typical progress curves for a segment, b) identifying students who are having difficulties, c) determining appropriate course and segment lengths for both classes and individuals, and d) conducting formal and informal evaluations of training effectiveness.

Rationale: The benefits and advantages of having data available for training decision making have been described in earlier sections. There is a large potential for cost reduction from knowing how long to train and from the capability to realistically evaluate how well training is working; this is likely to overshadow the additional resources involved in implementing the required mechanisms. Collection of performance and/or productivity data is routine in manufacturing settings, and is apparently justifiable on a cost basis in those environments.

2. MORE FLEXIBILITY IN TRAINING TIME AND SCHEDULING.

Develop specific mechanisms which enable the adjusting of training time for classes and individuals on the basis of performance-related indices. Training managers should have the flexibility to provide extra time, to rearrange schedules, and to vary the method and pacing of instruction as required to exercise quality control over the training product.

Rationale: Historical training data will be of limited utility unless scheduling can be varied to accommodate to differences among trainees in learning speed and patterns. Maintaining a time-based structure as the principal determinant of segment length prohibits the use of both general learning guidance and any specific indications that might be developed from historical data; the only possible use of such data would be in the "enrichment" of training for more able students.

Quality of output should be given emphasis equal to that of scheduling, but practical constraints in the training environment must be acknowledged and provided for. A predictable flow of student output must be available, both between segments and from training to the job. Predictability need not involve lockstep scheduling. It should be possible to develop new scheduling and assignment strategies which retain predictability and involve minimal disruption of student flow while also improving quality. This is an extremely complex issue, both technically and from a force management standpoint.

3. PROGRAMS FOR REFRESHER TRAINING

Develop formal programs for routine provision of update and refresher training. These would initially focus on critical skills for specialists in selected jobs, with a gradual transition to all major job components for all specialties.

Requirements for refresher training should be established by policy, and training should be provided on a regular basis, either in an operational setting or through consolidated facilities at a higher organizational level.

Rationale: It is well established that skills decay over time without continuing on-job practice or specific rehearsal and relearning. Those tasks performed routinely as an integral part of job duties will be retained without intervention. Some critical skills (emergency actions, combat-specific procedures) are rarely used under routine job conditions and may not be available when needed. These critical components should be periodically refreshed to maintain proficiency. Where appropriate, training equipment, particularly part-task devices, should be considered for standardization and economy of proficiency instruction.

4. RESEARCH AND DEVELOPMENT ON TASK AND SKILL DESCRIPTION SYSTEMS

Develop a basic, standard notational system for describing military task requirements and trainee and operator capabilities.

Rationale: There are difficulties in applying the products of research and development on training and education because there is no consistent method for describing the requirements of a job and the skills and abilities of the entrant and post-training populations. Numerous schemes have been proposed for such descriptions. Most writers, however, have felt it necessary to cover the full range of theoretical issues in suggested classifications and taxonomies. It may be feasible to develop less ambitious, basic but effective descriptive schemes for task components and skills in the more applied context of military-unique training. These should approach categorization primarily from the standpoint of training prescriptions rather than attempt full theoretical coverage of the processes

underlying learning. While the latter concern may be important, it will probably inhibit the discovery of a simplified set of categories. Classifications should be sufficiently eclectic in orientation to be usable across the spectrum of military training, and their development should be responsive to joint-service needs.

REFERENCES*

Adams, J. A. (1957). The relationship between certain measures of ability and the acquisition of a psychomotor response. Journal of General Psychology, 56, 121-134.

Adams, J. A. (1968). Response feedback and learning. Psychological Bulletin, 70, 486-504.

Adams, J. A. (1971). A closed-loop theory of motor learning. Journal of Motor Behavior, 3, 111-149.

Adams, J. A. (1985, in preparation). Historical review of research on the learning, retention and transfer of human motor skills. Draft monograph.

Adams, J. A. & Hufford, L. E. (1962). Contributions of a part-task trainer to the learning and relearning of a time-shared flight maneuver. Human Factors, 4, 159-170.

Alvares, K. M. & Hulin, C. L. (1972). Two explanations of temporal changes in ability-skill relationships: A literature review and theoretical analysis. Human Factors, 14, 295-308.

Alvares, K. M. & Hulin, C. L. (1973). An experimental evaluation of temporal decay in the prediction of performance. Organizational Behavior and Human Performance, 9, 169-185.

Ambler, R. K. & Lane, N. E. (1974). An advanced pipeline assignment system for naval flight officer students. NAMRL-1195. Pensacola, FL: Naval Aerospace Medical Research Laboratory.

* Ammons, R. B., Farr, R. G., Bloch, E., Neumann, E., Dey, M., Marion, R., & Ammons, C. H. (1958). Long-term retention of perceptual-motor skills. Journal of Experimental Psychology, 55, 318-328.

Anderson, J. R. (Ed.) (1981). Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89, No. 4, 369-406.

Annett, J. (1971). Acquisition of skill. British Medical Bulletin, 27, 266-271.

* Bahrick, H. P. (1964). Retention curves: Facts or artifacts? Psychological Bulletin, 61, 188-194.

* Indicates a reference not cited in text

Bahrick, H. P., Fitts, P. M., & Briggs, G. E. (1957). Learning curves - facts or artifacts? Psychological Bulletin, 54, 156-268.

Baldwin, R. D., Cliborn, R. E., & Foskett, R. J. (1976). The acquisition and retention of visual aircraft recognition skills. Report No. 76-A-4. Alexandria, VA: Army Research Institute.

Baloff, N. & Becker, S. W. (1967). On the futility of aggregating individual learning curves. Psychological Reports, 20, 183-191.

Batson, W. H. (1916). Acquisition of skill. Psychological Monographs, 21, No. 3.

Battig, W. F. (1966). Facilitation and interference. In E. A. Bilodeau (Ed.), Acquisition of skill. New York: Academic Press.

Battig, W. F. (1972). Intratask interference as a source of facilitation in transfer and retention. In R. F. Thompson & J. F. Voss (Eds.), Topics in learning and performance. New York: Academic Press.

Berbaum, K. S., Kennedy, R. S., Welch, R. B. & Brannan, J. R. (1985) Space adaptation syndrome: Reducing symptomatology through perceptual training. Final Report, Contract No. NAS 9-17278, NASA Johnson Space Center. Orlando, FL: Essex Corporation.

Berliner, C., Angell, D., & Shearer, J. W. (1964). Behaviors, measures and instruments for performance evaluation in simulated environments. Proceedings of the Symposium and Workshop on the Quantification of Human Performance, August, 1964, Albuquerque, NM.

* Bernstein, B. R. & Gonzalez, B. K. (1971). Learning, retention, and transfer. NAVTRADEVCEEN 68-C-0215-1 (Vol. I). Orlando, FL: Naval Training Device Center.

Bilodeau, E. A. & Bilodeau, I. McD. (1961). Motor skills learning. Annual Review of Psychology, 12, 243-280.

Bittner, A. C., Carter, R. C., Kennedy, R. S., Harbeson, M. M. & Krause, M. (1984). Performance Evaluation Tests for Environmental Research (PETER): Evaluation of 112 measures. Report No. NBDL-84R006. New Orleans, LA: Naval Biodynamics Laboratory.

Bloom, B. S. (1974). Time and learning. American Psychologist, 29, 682-688.

Bond, N. A., Jr. (1985). Yamada's remarkable keyboard. Office of Naval Research (Far East) Scientific Bulletin, 10(1), 37-46.

Bricton, C. A. & Burger, W. J. (1976). Transfer of training effectiveness: A-7E Night Carrier Landing Trainer (NCLT). NAVTRAEQUIPCEN N61339-74-C-0079-1. Orlando, FL: Naval Training Equipment Center.

* Briggs, G. E. & Naylor, J. C. (1962). Relative efficiency of several training methods as a function of transfer task complexity. Journal of Experimental Psychology, 64, 505-512.

Bruner, J. S. (1966). Toward a theory of instruction. New York: Norton.

Bryan, L. B. & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. Psychological Review, 6, 345-375.

Carter, R. & Woldstad, J. (1985). Repeated measurements of spatial ability with the Manikin test. Human Factors, 27, 209-219.

Cheng, P. W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. Psychological Review, 92, 414-423.

Christensen, J. M. (1982). Training for optimal performance in complex technical systems. Presented at Conference on "Human Reliability in Complex Technical Systems," Royal Swedish Academy of Engineering Sciences, Stockholm, Sweden.

Christensen, J. M. & Mills, R. B. (1967). What does the operator do in complex systems? Human Factors, 9, 329-340.

* Christiaansen, R. E. (1980). Prose memory: Forgetting rates for memory codes. Journal of Experimental Psychology: Human Learning and Memory, 6, No. 5, 611-619.

Conway, R. W. & Shultz, A. (1959). The manufacturing progress function. Journal of Industrial Engineering, 10 (1), 39-54.

Cormier, S. M. (1984). Transfer of training: An interpretive review. Technical Report 608. Alexandria, VA: Army Research Institute.

* Cotterman, J. E. & Wood, M. E. (1967). Retention of simulated lunar landing mission skills: A test of pilot reliability. AMRL-TR-66-222. Dayton, OH: Aerospace Medical Research Laboratory.

Cronbach, L. J. (1957). The disciplines of scientific psychology. American Psychologist, 12, 671-684.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. American Psychologist, 30, 116-127.

- Cronbach, L. J. & Snow, R. E. (1976). Aptitude and instructional methods. New York: Irvington.
- Cronholm, J. N. (1985). The optimization of training systems. In Proceedings of Symposium on Transfer of Training to Military Operational Systems, NATO Panel VIII, October 1985, Brussels, Belgium.
- Dansereau, D. F., Atkinson, T. R., Long, G. L., & McDonald, B. (1974). Learning strategies: A review and synthesis of the current literature. AFHRL-TR-74-70. Lowry AFB, CO: Air Force Human Resources Laboratory.
- DeJong, J. R. (1957). The effects of increasing skill on cycle time and its consequences for time standards. Ergonomics, 1, 51-60.
- Digman, J. M. (1959). Growth of a motor skill as a function of distribution of practice. Journal of Experimental Psychology, 57, 311-316.
- Dollard, J. A., Dixon, M. L. & McCann, P. H. (1980). Instruction and training management with computer technology: A pilot application. NPRDC TR 80-34. San Diego, CA: Navy Personnel Research and Development Center.
- Eddowes, E. E. (1974). A cognitive model of what is learned during flying training. AFHRL-TR-74-63. Williams AFB, AZ: Air Force Human Resources Laboratory.
- Ehrlich, G. (1943). A method of constructing learning curves for a motor skill involving total body speed and accuracy. Journal of Applied Psychology, 28, 494-503.
- Estes, W. K. (1956). The problem of inferences based on group data. Psychological Bulletin, 53, 134-140.
- Farr, M. J. (1986). The long-term retention of knowledge and skills: A cognitive and instructional perspective. M-205. Alexandria, VA: Institute for Defense Analyses.
- Ferguson, G. A. (1956). On transfer and the abilities of man. Canadian Journal of Psychology, 10, 121-131.
- Fischer, G., Brown, J. S. & Burton, R. R. (1978). Aspects of a theory of simplification, debugging and coaching. BBN Report No. 3912. Cambridge, MA: Bolt, Beranek & Newman.
- Fitts, P. M. (1962). Factors in complex skill training. In R. Glaser (Ed.), Training research and education. Pittsburgh: Pittsburgh University Press.
- Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), Categories of human learning. New York, NY: Academic Press.

Fitts, P. M. & Posner, M. I. (1967). Human performance. Monterey, CA: Brooks/Cole.

Fitts, P. M. & Seeger, C. M. (1953). SR compatibility: Spatial characteristics of stimulus and response codes. Journal of Experimental Psychology, 56, 78-81.

Fleishman, E. A. (1966). Human abilities and the acquisition of skill. In E. A. Bilodeau (Ed.), Acquisition of skill. New York, NY: Academic Press.

Fleishman, E. A. (1972). On the relation between abilities, learning, and human performance. American Psychologist, 27, 1017-1032.

Fleishman, E. A. (1982). Systems for describing human tasks. American Psychologist, 37, 821-834.

* Fleishman, E. A. & Parker, S. F. (1962). Factors in the retention and relearning of perceptual motor skill. Journal of Experimental Psychology, 64, 215-276.

Fleishman, E. A. & Quaintance, M. K. (1984). Taxonomies of human performance: The description of human tasks. Orlando, FL: Academic Press.

Frederickson, C. H. (1969). Abilities, transfer and information retrieval in verbal learning. Multivariate Behavioral Research Monographs, No. 69-2.

Gagne, R. M. (1977). The conditions of learning. (3rd Ed.). New York: Holt, Rinehart & Winston.

Gagne, R. M. & Briggs, L. J. (1979). Principles of instructional design (2nd ed.). New York: Holt, Rinehart & Winston.

Gagne, R. M. & Dick, W. (1983). Instructional psychology. Annual Review of Psychology, 34, 261-295.

Gardlin, G. R. & Sitterley, T. E. (1972). Degradation of learned skills - A review and annotated bibliography. Report D180-15080-1. Seattle, WA: Boeing.

Gentner, D. R. (1983). The acquisition of typewriting skill. Acta Psychologica, 54, 233-248.

Gettinger, M. (1984). Individual differences in time needed for learning: A review of literature. Educational Psychologist, 19, 25-29.

Gettinger, M. & White, M. A. (1979). Which is the stronger correlate of school learning? Time to learn or measured intelligence? Journal of Educational Psychology, 71, 405-412.

Glaser, R. (1962). Introduction: Psychology and instructional technology. In R. Glaser (Ed.), Training research and education. Pittsburgh: Pittsburgh University Press.

Glaser, R. (1976). Components of a psychology of instruction: Toward a science of design. Review of Educational Research, 46, 1-24.

Glaser, R. (1982). Instructional psychology: Past, present and future. American Psychologist, 37 (3), 292-305.

Glaser, R. & Resnick, L. B. (1972). Instructional psychology. Annual Review of Psychology, 23, 181-276.

Glover, J. H. (1966). Manufacturing progress functions: II. Selection of trainees and control of their progress. International Journal of Production Research, 5(1), 43-59.

Gopher, D., Koenig, W., Karis, D., & Donchin, E. (1984). An information processing approach to the study of data entry skills: The effects of representation rules and coordination requirements. Technical Report CPL-84-4. Champaign, IL: Cognitive Psychophysiology Laboratory, University of Illinois.

* Green, B. F. & Hall, J. A. (1984). Quantitative methods for literature reviews. Annual Review of Psychology, 35, 37-53.

* Grimsley, D. L. (1969). Acquisition, retention and retraining: Effects of high and low fidelity in training devices. Technical Report 69-1. Alexandria, VA: Human Resources Research Organization.

Hackett, E. A. (1974). CNAAM Phil. Thesis, Middlesex Polytechnic at Hendon. Cited in: Towill, D. R. (1976). Transfer functions and learning curves. Ergonomics, 19, 623-638.

* Hagman, J. D. (1980). Effects of training task repetition on retention and transfer of maintenance skill. Report 1271. Alexandria, VA: Army Research Institute.

Hagman, J. D. & Rose, A. M. (1983). Retention of military tasks: A review. Human Factors, 25, 199-213.

* Hammerton, M. (1963). Retention of learning in a difficult tracking task. Journal of Experimental Psychology, 66, 108-110.

Hancock, W. M. (1971). The learning curve. In H. B. Maynard (Ed.), Industrial Engineering Handbook (3rd ed.). New York: McGraw-Hill.

Hayes, K. J. (1953). The backward curve: A method for the study of learning. Psychological Review, 60, 269-275.

Hayes, K. J. & Pereboom, A. C. (1959). Artifacts in criterion-reference learning curves. Psychological Review, 66(1), 23-26.

Horton, D. L. & Mills, C. B. (1984). Human learning and memory. Annual Review of Psychology, 35, 361-394.

* Hudson, P. R. (1984). Knowledge and skill retention: A bibliography. The Technical Cooperation Program (TTCP), Technical Panel UTP-2 on Training Technology.

Hull, C. L. (1943). Principles of behavior. New York: Appleton Century Crofts.

Hull, C. L. (1952). Essentials of behavior. New Haven: Yale University Press.

Humphreys, L. G. (1960). Investigation of the simplex. Psychometrika, 25, 313-323.

Irion, A. L. (1966). A brief history of research on the acquisition of skill. In E. A. Bilodeau (Ed.), Acquisition of skill. New York: Academic Press.

* Isley, R. N. & Spears, W. D. (1982) Phase I pilot study: VTRS transfer of training experiment. NAVTRAEQUIPCEN N61339-80-D-0009-17-2. Orlando, FL: Naval Training Equipment Center.

Isley, R. N., Spears, W. D., Prophet, W. W., & Corley, W. E. (1982). VTRS transfer-of-training experiment: Phase I - An experimental design. NAVTRAEQUIPCEN N61339-80-D-009-17-1. Orlando, FL: Naval Training Equipment Center.

* Johnson, K. A. (1969). Retention of electronic fundamentals: Differences among topics. Technical Bulletin STB 70-1. San Diego, CA: Navy Personnel Research and Development Center.

Johnson, S. L. (1980). A cost analysis method of establishing trainer criteria. Ergonomics, 23, 1137-1145.

Jones, M. B. (1959). Simplex theory. Monograph No. 3. Pensacola, FL: Naval School of Aviation Medicine.

Jones, M. B. (1960). Molar correlational analysis. Monograph No. 4. Pensacola, FL: Naval School of Aviation Medicine.

Jones, M. B. (1969). Differential processes in acquisition. In E. A. Bilodeau (Ed.), Principles of skill acquisition. New York: Academic Press.

Jones, M. B. (1970). A two-process theory of individual differences in motor learning. Psychological Review, 77, 353-360.

Jones, M. B. (1985). Nonimposed overpractice and skill retention. Final report, Contract MDA 903-83-K-0246 for Army Research Institute.

Jones, S. A. & Kuntz, L.-A. (1985). The relative efficiencies of whole, part or impoverished-part task training on a complex perceptual-motor skill. Unpublished practicum manuscript. Orlando, FL: University of Central Florida.

Kao, D. L. (1937). Plateaus and the curve of learning in motor skill. Psychological Monographs, 49, No. 3, Whole No. 219.

Keller, F. S. (1958). The phantom plateau. Journal of Experimental Analysis of Behavior, 1, 1-13.

Kennedy, R. S., Dunlap, W. P., Wilkes, R. L. & Lane, N. E. (1985). Factor analysis of a fifteen-minute microcomputer-based performance test battery. EOTR 85-2. Orlando, FL: Essex Corporation.

Kennedy, R. S., Wilkes, R. L., Lane, N. E., & Homick, J. L. (1985). Preliminary evaluation of a micro-based repeated measures testing system. EOTR-85-1. Orlando, FL: Essex Corporation.

Kieras, D. E. (1985). The role of prior knowledge in operating equipment from written instructions. Report No. 19 (FR-85/ONR-19). Ann Arbor, MI: University of Michigan.

Kieras, D. E. & Bovair, S. (1985). The acquisition of procedures from text: A production-system analysis of transfer of training. Technical Report TR-85/ONR-16. Ann Arbor, MI: University of Michigan.

Kinkade, R. G. (1963). A differential influence of augmented feedback on learning and on performance. AMRL-TDR-63-12. Dayton, OH: Aerospace Medical Research Laboratory.

Kleinman, M. (1977). Ability factors in motor learning. Perceptual and Motor Skills, 44, 827-836.

Knerr, C. M. & Sticha, P. J. (1985). Models of learning and performance of armor skills. In Proceedings of Symposium on Transfer of Training to Military Operational Systems, NATO Panel VIII, October 1985, Brussels, Belgium.

Kolers, P. A. & Duchnick, R. L. (1985). Discontinuity in cognitive skill. Journal of Experimental Psychology: Learning, Memory and Cognition, 11, 655-674.

Krendel, E. S. & Bloom, J. W. (1963). The natural pilot model for flight proficiency evaluation. NAVTRADEVCEEN 323-1. Port Washington, NY: Naval Training Device Center.

Krueger, W. C. F. (1947). Influence of difficulty of perceptual motor task upon acceleration of curves of learning. Journal of Educational Psychology, 38, 51-53.

Kulhavy, R. W. (1977). Feedback in written instruction. Review of Educational Research, 47, 211-232.

Lane, N. E. & Ambler, R. K. (1974). Impact of multivariate analysis on the aviation selection and classification process. In Medical requirements and examination procedures in relation to tasks of today's aircrew. Conference Proceedings CP-153. Neuilly-sur-Seine, France: North Atlantic Treaty Organization.

Lee, T. D. & Magill, R. A. (1983). The locus of contextual interference in motor skill acquisition. Journal of Experimental Psychology: Learning, Memory and Cognition, 9, 730-746.

Leonard, R. L., Wheaton, G. R., & Cohen, F. D. (1976). Transfer of training and skill retention. TR-76-A3. Alexandria, VA: Army Research Institute.

Levy, F. K. (1965). Adaptation in production processes. Management Science, II(6), B136-B154.

Lewis, D. (1949). Quantitative methods in psychology. Ann Arbor, MI: Edwards.

Lintern, G. (1985)(a). A perceptual learning approach to skill transfer for manual control. NAVTRAEQUIPCEN 81-C-0105-12. Orlando, FL: Naval Training Equipment Center.

Lintern, G. (1985)(b). Context and dependency effects in skill acquisition. In R. E. Eberts & C. G. Eberts (Eds.), Trends in ergonomics/human factors II. Amsterdam: Elsevier Science Publishers.

Lintern, G. & Roscoe, S. N. (1980). Visual cue augmentation in contact flight simulation. In S. N. Roscoe (Ed.), Aviation psychology. Ames, IA: Iowa State University Press.

Lintern, G., Thomley, K. E., Nelson, B. E., & Roscoe, S. N. (1984). Content, variety, and augmentation of simulated visual scenes for teaching air-to-ground attack. NAVTRAEQUIPCEN 81-C-0105-3. Orlando, FL: Naval Training Equipment Center.

Mane, A. M. (1984). Acquisition of perceptual-motor skill: Adaptive and part-whole training. In Proceedings of the Human Factors Society: 28th Annual Meeting, 522-526.

Mane, A. M., Coles, M. G., Wickens, C. D. & Donchin, E. D. (1983). The use of the additive factors methodology in the analysis of skill. In Proceedings of the Human Factors Society: 27th Annual Meeting, 407-411.

* Marcus, A. J. & Curran, L. E. (1985). The use of flight simulators in measuring and improving training effectiveness. In Proceedings of Symposium on Transfer of Training to Military Operational Systems, NATO Panel VIII, October 1985, Brussels, Belgium.

Martin, E. L. (1984). Practice makes perfect. AFHRL-TP-84-32. Williams AFB, AZ: Air Force Human Resources Laboratory.

Martin, E. L. & Waag, W. L. (1978). Contribution of platform motion to simulator training effectiveness: Study I: Basic contact. AFHRL-TR-78-15. Williams AFB, AZ: Air Force Human Resources Laboratory.

Mazur, J. E. & Hastie, R. (1978). Learning as accumulation: A reexamination of the learning curve. Psychological Bulletin, 85, 1256- 1274.

McCormick, E. J. & Jeanneret, P. R. (1984). The Position Analysis Questionnaire (PAQ). In S. Gael (Ed.), Job Analysis Handbook. New York: Wiley.

* McDonald, R. D. (1967). Retention of military skills acquired in basic combat training. Technical Report 67-13. Alexandria, VA: Human Resources Research Office.

McKeachie, W. J. (1974). Instructional psychology. Annual Review of Psychology, 25, 161-193.

* Melnick, M. J. (1971). Effects of overlearning on the retention of a gross motor skill. The Research Quarterly, 42, No. 1, 62-69.

* Mengelkoch, R. F., Adams, J. A., & Gainer, C. (1971). The forgetting of instrument flying skills. Human Factors, 13, 397-405.

Miller, R. B. (1962). Analysis and specification of behavior for training. In R. Glaser (Ed.), Training research and education. Pittsburgh, PA: University of Pittsburgh Press.

Miller, R. B. (1975). Taxonomies for training. In W. T. Singleton & P. Spurgeon (Eds.), Measurement of human resources. London: Francis & Taylor.

Morrison, J. E. (1982). Retention of armor procedures: A structural analysis. Technical Report 591. Alexandria, VA: Army Research Institute.

Nadler, G. & Smith, W. D. (1963). Manufacturing progress functions for types of processes. International Journal of Production Research, 2(2), 115-135.

Nanda, R. (1977). Learning curves: An overview. In R. Nanda & G. L. Adler (Eds.), Learning curves: Theory and application. Atlanta, GA: American Institute of Industrial Engineers.

Nanda, R. & Adler, G. L. (Eds.) (1977). Learning curves: Theory and application. Atlanta, GA: American Institute of Industrial Engineers.

Naylor, J. C. (1962). Factors affecting the relative efficiency of part- and whole-practice methods: A review of the literature. Technical Report 950-1. Port Washington, NY: Naval Training Devices Center.

Naylor, J. C. & Briggs, G. E. (1961). Long-term retention of learned skills: A review of the literature. ASD-TR-61-390. Dayton, OH: Aerospace Medical Research Laboratory.

Neves, D. M. & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Newell, A. & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Noble, C. E., Salazar, O. G., Skelley, C. S., & Wilkerson, H. R. (1979). Work and rest variables in the acquisition of psychomotor tracking skill. Journal of Motor Behavior, 11, 233-246.

Norman, D., Gentner, D., Stevens, A. (1976). Comments on learning schemata and memory representation. In D. Klahr (Ed.), Cognition and instruction. Hillsdale, NJ: Erlbaum.

Orlansky, J. & String, J. (1981). Cost-effectiveness of maintenance simulators for military training. IDA Paper P-1568. Alexandria, VA: Institute for Defense Analyses.

Pegels, C. C. (1969). On startup or learning curves: An expanded view. AIEE Transactions, 1(3), 216-222.

Peterson, F. E., Booth, R. F., Lane, N. E. & Ambler, R. K. (1967). Predicting success in naval flight officer training. NAMI-996. Pensacola, FL: Naval Aerospace Medical Institute.

Powers, D. E. (1976). Instructional strategies and individual differences: A selective review and summary of literature. Report No. 4 on DARPA Contract MDA903-74-C-0290. Princeton, NJ: Educational Testing Service.

Prophet, W. W. (1976). Long-term retention of flying skills: A review of the literature. Technical Report 76-35. Alexandria, VA: Human Resources Research Organization.

Rasmussen, J. (1979). On the structure of knowledge: A morphology of mental models in man-machine systems context. Report M-2192. Riso National Laboratory, Denmark.

Reigeluth, C. M., Doughty, P., Sari, F., Powell, C. J., Frey, L., & Sweeney, J. (1982). Instructional design literature review for EDeP (Extended Design Procedure). Final Report Contract TNC81-115(N) RO 741-81. Fort Monroe, VA: US Army Training and Doctrine Command.

Resnick, L. B. (1981). Instructional psychology. Annual Review of Psychology, 32, 659-704.

Reynolds, B. (1952)(a). Correlation between two psychomotor tasks as a function of distribution of practice on the first. Journal of Experimental Psychology, 43, 341-348.

Reynolds, B. (1952)(b). The effect of learning on the predictability of psychomotor performance. Journal of Experimental Psychology, 44, 189-198.

Ricard, G. L. & McWilliams, E. (1985). Intelligent simulation for part-task training. Presented at NATO Conference on "Computer-Based Instruction in Military Environments," April 1985, Brussels, Belgium (to be published by North Holland Press).

Rose, A. M., Czarnolewski, M. Y., Gragg, F. E., Austin, S. H., Ford, P., Doyle, J., & Hagman, J. D. (1984). Acquisition and retention of soldiering skills. Report No. AIR FR88600. Washington, DC: American Institute for Research.

Rowland, D. (1985). Assessment of the effects of experience on combat effectiveness. In Proceedings of Symposium on the Military Value and Cost-Effectiveness of Training, NATO PANEL VII, January 1985, Brussels, Belgium.

Rumelhart, D. E. & Norman, D. A. (1978). Accretion, tuning and restructuring: Three modes of learning. In J. W. Cotton & R. L. Klatzky (Eds.), Semantic factors in cognition. Hillsdale, NJ: Erlbaum.

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical appraisal. Psychological Bulletin, 95, 355-386.

Savage, R. E., Williges, B. H. & Williges, R. C. (1982). Empirical prediction models for training-group assignment. HUMAN Factors, 24, 417-426.

* Schendel, J. D. & Hagman, J. D. (1982). On sustaining procedural skills over a prolonged retention interval. Journal of Applied Psychology, 67, 605-610.

Schendel, J. D., Shields, J. L., & Katz, M. S. (1978). Retention of motor skills: Review. Technical Paper 313. Alexandria, VA: Army Research Institute.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. Psychological Review, 82, 225-260.

Schmidt, R. A. (1982). The schema concept. In J. A. Kelso (Ed.), Human motor behavior: An introduction. Hillsdale, NJ: Erlbaum.

Schneider, W. (1984). Toward a model of attention and the development of automatic processing. HARL-ONR-8402. Champaign, IL: Human Attention Research Laboratory, University of Illinois.

Schneider, W. (1985). Training high performance skills: Fallacies and guidelines. Human Factors, 27, 285-300.

Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. Psychological Review, 84, 1-66.

Schneider, W. & Shiffrin, R. (1985). Categorization (restructuring) and automatization: Two separable factors. Psychological Review, 92, 424-428.

Shapiro, D. C. & Schmidt, R. A. (1982). Development of schemes, plans and programs. In J. A. Kelso & J. E. Clark (Eds.), The development of movement control and coordination. New York: Wiley.

Shea, J. F. & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention and transfer of a motor skill. Journal of Experimental Psychology: Human Learning and Memory, 5, 179-187.

Shoenberger, R. W., Wherry, R. J., Jr. & Berkshire, J. R. (1963). Predicting success in aviation training. NSAM-873. Pensacola, FL: Naval School of Aviation Medicine.

Shephard, A. H. & Lewis, D. (1950). Prior learning as a factor in shaping performance curves. SDC938-1-4. Port Washington, NY: Navy Special Devices Center.

Sheppard, D. J. (1985). Visual and part-task manipulations for teaching simulated carrier landings. NAVTRAEQUIPCEN 81-C-0105-9. Orlando, FL: Naval Training Equipment Center.

* Shields, J. L., Goldberg, S. L., & Dressel, J. D. (1979). Retention of basic soldiering skills. Research Report No. 1225. Alexandria, VA: Army Research Institute.

Shiffrin, R. M. & Dumais, S. T. (1981). The development of automatism. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

* Shuell, T. J. & Keppel, G. (1970). Learning ability and retention. Journal of Educational Psychology, 61, 59-65.

Singer, N. & Gerson, R. F. (1978). Cognitive processes and learner strategies in the acquisition of motor skills. Report No. TR-78-TH-10. Alexandria, VA: Army Research Institute.

* Sitterley, T. E. (1974). Degradation of learned skills: Static practice effectiveness for visual approach and landing skill retention. Report D180- 17876-1. Seattle, WA: Boeing Aerospace.

Sitterley, T. E. & Berge, W. A. (1972). Degradation of learned skills: Effectiveness of practice methods on simulated space flight skill retention. Report D180-15081-1. Seattle, WA: Boeing Aerospace.

Slamecka, N. J. & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 384-397.

Smith, J. F. (1976). Continuation versus recurrent pilot training. AFHRL-TR-76-4. Williams AFB, AZ: Air Force Human Resources Laboratory.

Snoddy, G. S. (1926). Learning and stability. Applied Psychology, 10, 1-36.

Sparrow, J., Patrick, J., Spurgeon, P., & Barwell, F. (1982). The use of job component analysis and related aptitudes in personnel selection. Journal of Occupational Psychology, 55, 157-164.

Spears, W. D. (1985). Measurement of learning and transfer through curve fitting. Human Factors, 27, 251-266.

Spears, W. D. (1983). Processes of skill performance: A foundation for the design and use of training equipment. NAVTRAEQUIPCEN 78-C-0013-4. Orlando, FL: Naval Training Equipment Center.

Stevens, J. C. & Savin, H. B. (1962). On the form of learning curves. Journal of the Experimental Analysis of Behavior, 5, 15-18.

Swope, W. M., Corey, J. M., Evans, R. M. & Morris, C. L. (1982). Analysis of factors affecting the performance of the Navy's computer-managed instructional system. Technical Report 119. Orlando, FL: Training Analysis and Evaluation Group.

Taylor, J. E., Montague, E. K., & Hauke, R. (1970). The interrelationships of ability level, instructional system, and skill acquisition. Professional Paper 29-70. Presidio of Monterey, CA: Human Resources Research Organization.

Taylor, J. E. & Thalman, D. M. (1977). An examination of skill deterioration and retraining in the United States Navy. Unpublished master's thesis. Monterey, CA: Naval Postgraduate School.

Taylor, J. G. & Smith, P. C. (1956). An investigation of the shape of learning curves for industrial motor tasks. Journal of Applied Psychology, 40, 142-149.

Thurstone, L. L. (1919) The learning curve equation. Psychological Monographs, 26, No. 114.

Towill, D. R. (1976). Transfer functions and learning curves. Ergonomics, 19, 623-638.

Underwood, B. J. (1964). Degree of learning and the measurement of forgetting. Journal of Verbal Learning and Verbal Behavior, 3, 112-129.

Valverde, H. H. (1973). A review of flight simulator transfer-of- training studies. Human Factors, 15, 510-523.

Vanderplas, J. M. (1958). Transfer of training and its relation to perceptual learning and recognition. Psychological Review, 65, 375-385.

* Vineberg, R. (1975). A study of the retention of skills and knowledge acquired in basic training. HUMRRO-TR-75-10. Alexandria, VA: Human Resources Research Organization.

Wexley, K. N. (1984). Personnel training. Annual Review of Psychology, 35, 519-551.

Wheaton, G. R., Rose, A. M., Fingerman, P. W., Korotkin, A. L., & Holding, D. H. (1976). Evaluation of the effectiveness of training devices: Literature review and preliminary model. Research Memo 76-6. Alexandria, VA: Army Research Institute.

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. Memory and Cognition, 2, 775-780.

Wightman, D. C. (1983) Part-task training strategies in simulated carrier landing final approach training. NAVTRAEQUIPCEN IH-347. Orlando, FL: Naval Training Equipment Center.

Wightman, D. C. & Lintern, G. (1985). Part-task training for tracking and manual control. Human Factors, 27, 267-283.

Williges, B. H. & Williges, R. C. (1980). Automated motor skills training optimized for individual differences. AFOSR-TR-81-0077. Blacksburg, VA: Virginia Polytechnic and State University.

Wisher, R. (1985). The design and evaluation of hand-held training devices (CHIP). Presented at NATO Conference on Computer-Based Instruction in Military Environments, April 1985, Brussels, Belgium (to be published by North Holland Press).

Woodrow, H. (1938). The relation between abilities and improvement with practice. Journal of Educational Psychology, 29, 215-230.

Woodrow, H. (1940). Interrelations of measures of learning. Journal of Psychology, 10, 49-73.

Woodrow, H. (1946). The ability to learn. Psychological Review, 53, 147-158.

Wright, T. P. (1936) Factors affecting the cost of airplanes. Journal of the Aeronautical Sciences, 3(4), 122-128.

Wrisberg, C. A. & Winter, T. P. (1983). Variability of practice and the transfer of training of motor skills. Technical Report 596. Alexandria, VA: Army Research Institute.

Yen, W. M. (1978). Measuring individual differences with an information-processing model. Journal of Educational Psychology, 70, 72-86.